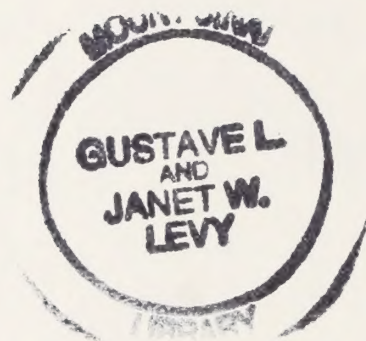
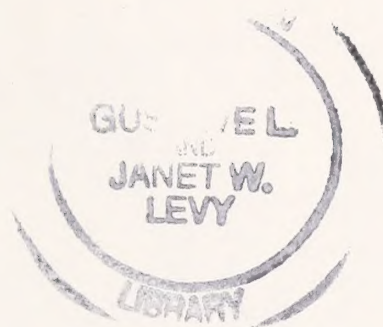
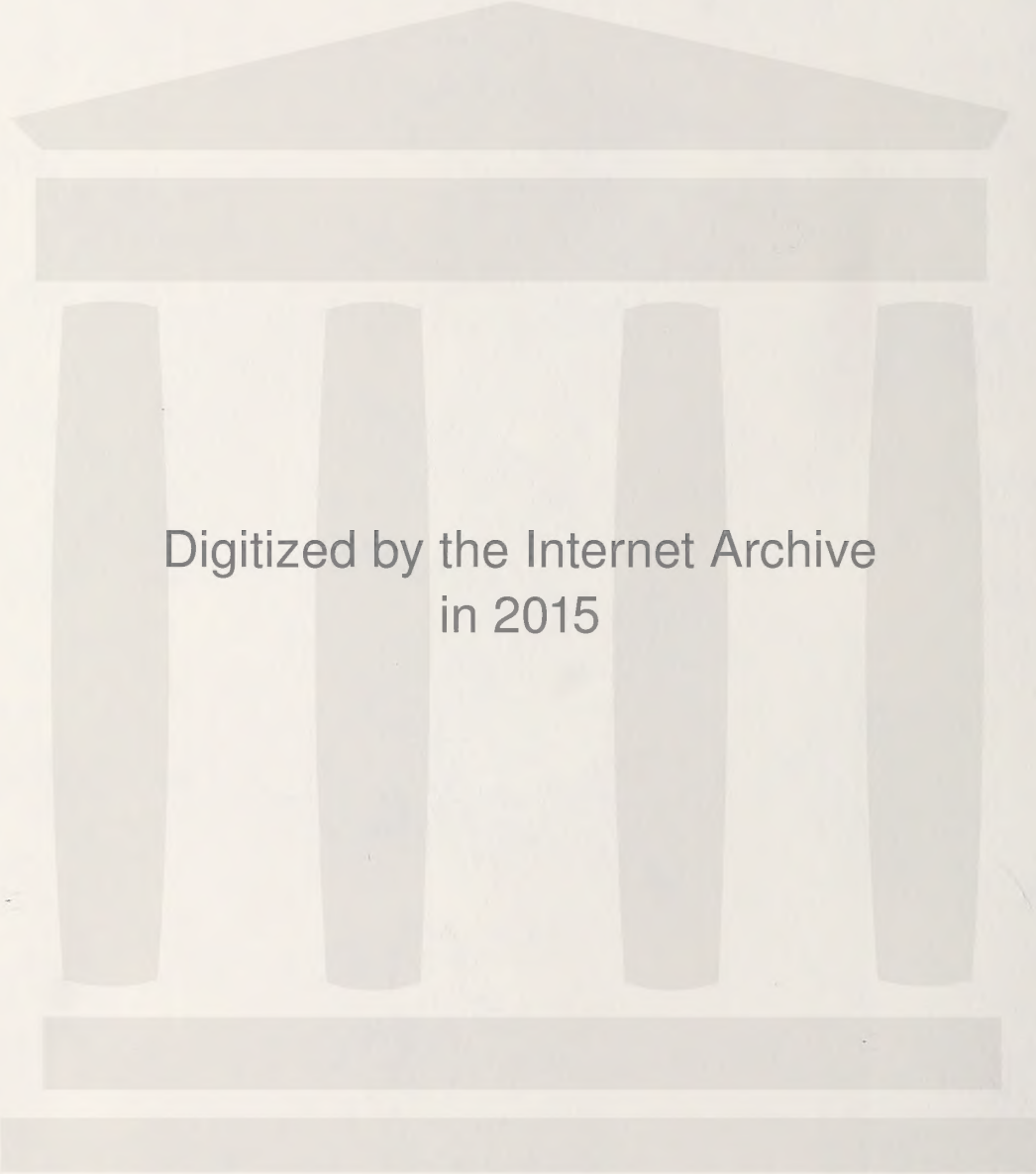








Digitized by the Internet Archive
in 2015

<https://archive.org/details/mountsinaijourna7620moun>

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 76, NO. 1
JANUARY/FEBRUARY 2009

Issue Theme
Advanced Lung Disease:
Comorbidities and
Challenges

Journal of Translational and Personalized Medicine



Spectrum of Fibrosing Diffuse Parenchymal Lung Disease
Gastroesophageal Reflux Disease and Idiopathic Pulmonary Fibrosis
"End Stage" Pulmonary Fibrosis in Sarcoidosis
High Resolution Computed Tomography in Idiopathic Interstitial Pneumonias
Comorbidities of Advanced Lung Disease
Advanced Lung Disease: Quality of Life and Role of Palliative Care

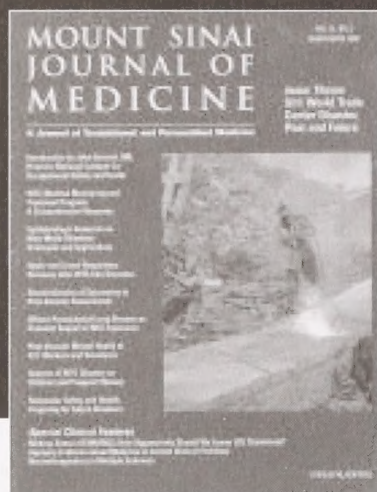
Special Features

Healthcare-associated Infections: Epidemiology, Prevention, and Therapy
Experience with Fistula in Niger: An Ethical Dilemma
Cetominophen Hepatotoxicity: Current Management

 **WILEY-BLACKWELL**

Newly
Relaunched!

Keep Abreast of the Latest Trends in
Translational and Personalized Medicine...
Enter Your Subscription Today!



MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Editor-in-Chief:

Penny A. Asbell, MD, FACS, MBA

Mount Sinai School of Medicine, New York, NY

2008 Themes and Guest Editors

Quality of Healthcare and Racial
Health Disparities (Mark Chassin)

9/11 World Trade Center Disaster:
Past and Future (Robin Herbert,
Jacqueline Moline)

Anxiety and Depression:
Leading Edge of Therapy
(Dennis Charney, Mathew Sanjay)

Diabetes Type 1:
New Light on Etiology,
New Approaches to Management
(Derek Leroith)

Hospital Medicine:
An Ascending New Field
(Andrew Dunn)

Advanced Lung Disease:
Meeting the Challenges
(Maria Padilla)

Continuously published since 1934 and included in *Index Medicus* from the very first volume, the **Mount Sinai Journal of Medicine (MSJM)** has a distinguished career as a general medical journal for a worldwide audience.

Building upon this tradition of education across medical specialties, **MSJM** was re-launched in 2007 as a journal of translational and personalized medicine, two disciplines in the forefront of modern healthcare. **MSJM** features articles on the evolving nature of clinical care, its close linkage to bio-medical research and the resulting opportunities for enhanced patient care.

Dedicated to improving communication among all participants in medicine today, **MSJM** is an excellent source of up-to-date medical information for healthcare clinicians, educators, researchers and policy makers worldwide. **MSJM** will inspire the creative thinking needed to create new ways to understand, prevent, and treat human disease.

Each issue includes several quality, in-depth review articles related to a single theme and several others on subjects of diverse clinical interest. All articles are written by renowned authorities in their fields, from both Mount Sinai and elsewhere, under the supervision of Guest Editors.



WILEY-BLACKWELL

Enter Your 2009 Subscription Today!



Your 2009 subscription will include the six 2009 print issues of the **Mount Sinai Journal of Medicine** (Volume 76), plus concurrent full-text online access to these six issues as they are posted on Wiley's Internet publishing platform.

Two Ways for Individuals to Subscribe:

1. Mount Sinai Alumna/Alumnus:
\$40.00 (Worldwide)
2. Personal Subscription:
\$95.00 (U.S./Canada/Mexico)
\$120.00 (Outside North America)

Complimentary Online Access for Institutions!

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

PERSONAL SUBSCRIPTION ORDER FORM

☐ Yes, please enter my 2009 personal subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$95.00 (U.S./Canada/Mexico) ☐ \$120.00 (Outside North America)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

Method of Payment (Canada residents please add GST. Personal rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **John Wiley & Sons, Inc.** Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA

Card # _____ Exp. Date _____

Signature _____
(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM0909

MSJM MEMBER SUBSCRIPTION ORDER FORM

☐ Yes, please enter my 2009 Alumna/Alumnus Subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$40.00 (Worldwide)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

METHOD OF PAYMENT (Alumna/Alumnus rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **The Mount Sinai School of Medicine**. Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA ☐ Other

Card # _____ Exp. Date _____

Signature _____
(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM09

COMPLIMENTARY ONLINE ACCESS FOR INSTITUTIONS

Mount Sinai Journal of Medicine

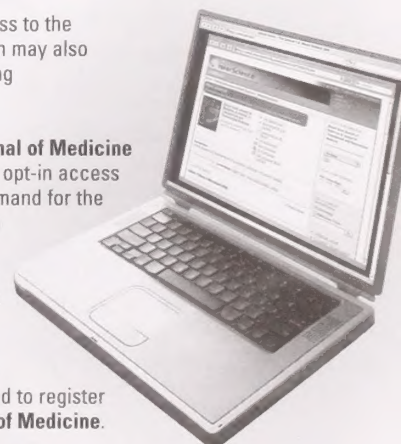
A Journal of Translational and Personalized Medicine

2009. Volume 76. 6 issues per year. Print ISSN: 0027-2507. Online ISSN: 1931-7581.

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

Institutional online and print pricing for the **Mount Sinai Journal of Medicine** will be set in 2009, subsequent to introductory complimentary opt-in access during 2008 and 2009. Pricing will be based on the overall demand for the Journal and on the size and type of the subscribing institution (e.g. academic, government, hospital, or corporate). Institutions that receive complimentary online access during 2008 and 2009 will have archival access in subsequent years according to Wiley's licensing terms and conditions.

Ask your librarian to e-mail optinaccess@wiley.co.uk or visit www.interscience.wiley.com/msjregister for more details and to register for complimentary online access to the **Mount Sinai Journal of Medicine**.





NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

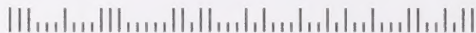
FIRST CLASS MAIL

PERMIT NO. 2277

HOBOKEN, NJ

POSTAGE WILL BE PAID BY ADDRESSEE

JOHN WILEY & SONS INC
ATTN: PETER JANSIC, MAIL STOP 8-02
111 RIVER STREET
HOBOKEN, NJ 07030-9441



PLACE
POSTAGE
STAMP
HERE

INES SANTOS
MOUNT SINAI JOURNAL OF MEDICINE
ONE GUSTAV L. LEVY PLACE BOX 1094
NEW YORK, NY 10029-6574

**Library
Routing
Card**

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis S. Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner
Thomas M. Coffman
Dwight Evans
Gabriel Haddad
Heidi E. Hamm
Matti Häyry
Eve J. Higginbotham
Katherine A. High
Tony Hope
John H. Krystal
Raju Kucherlapati
Robert J. Lefkowitz
Robert Lisak
George Lister
Philip A. Marsden
Charles Nemeroff
Scott Rauch
Bernard Roizman
Alan F. Schatzberg
Inder Verma
Jeffrey A. Whitsett
Allan Wolkoff
Noam Zohar

University of Washington
Duke University
University of Pennsylvania
University of California, San Diego
Vanderbilt University
University of Manchester
Morehouse School of Medicine
University of Pennsylvania
University of Oxford
Yale University
Harvard Medical School
Duke University
Wayne State University
University of Texas
University of Toronto
Emory University
Harvard Medical School
University of Chicago
Stanford University
The Salk Institute
University of Cincinnati
Albert Einstein College of Medicine
Bar-Ilan University

Theme Editors

Maria L. Padilla, MD
Director, Advanced Lung Disease/
Interstitial Lung Disease Program
Mount Sinai School of Medicine

Virginia R. Litle, MD
Associate Professor of Surgery
Division of Thoracic and Foregut Surgery
University of Rochester Medical Center

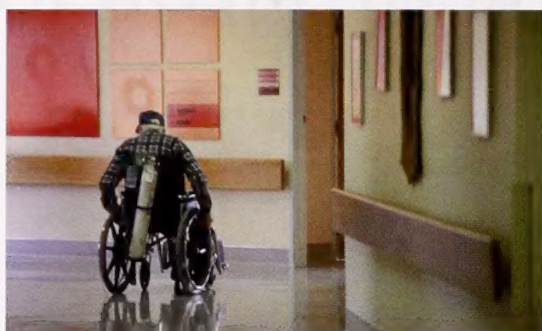
Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com

Advanced Lung Disease: Comorbidities and Challenges



Themes for future issues of *MSJM*, each shown with Guest Editor(s)

- Traumatic brain injury (Andrew Jagoda)
- Zoonotic infections (Stephanie Factor and Mary Klotman)
- Alzheimer's disease (Nikos Robakis)
- Type 2 diabetes mellitus (Dina Green)
- Drug discovery in academic setting (Ravi Iyengar)
- Personalized medicine (Erwin Bottinger)
- Vascular Surgery (Michael Marin)

©2009 Mount Sinai School of Medicine. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: \$95.00 in North America, \$120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinajournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS

VOL 76, NO 1

Theme Introduction

- 1** Advanced Lung Disease: Meeting the Challenges
Maria L. Padilla

Theme Articles

- 2** Spectrum of Fibrosing Diffuse Parenchymal Lung Disease
Adam S. Morgenthau and Maria L. Padilla
- 24** Gastroesophageal Reflux Disease and Idiopathic Pulmonary Fibrosis
Yevgenia Y. Pashinsky, Barry W. Jaffin, and Virginia R. Little
- 30** "End Stage" Pulmonary Fibrosis in Sarcoidosis
Alvin Teirstein and Adam Morgenthau
- 37** High Resolution Computed Tomography in Idiopathic Interstitial Pneumonias
Pamela A. Argiriadi and David S. Mendelson
- 53** Comorbidities of Advanced Lung Disease
Joel Anthony Nations and Steven D. Nathan
- 63** Advanced Lung Disease: Quality of Life and Role of Palliative Care
Christopher R. Gilbert and Cecilia M. Smith

Feature Department

- 71** Ethics [Editor: Rosamond Rhodes]
Fistula Experience in Niger: How We Overcame an Ethical Dilemma
Cedric K. Olivera, Charls J. Ascher-Walsh and Nada Gligorov

Special Features

- 75** Acetaminophen Hepatotoxicity: Current Management
Hui-Hui Tan, Charissa Y. Chang, and Paul Martin
- 84** Healthcare-associated Infections: Epidemiology, Prevention, and Therapy
Rupali K. Doshi, Gopi Patel, Richard MacKay, and Frances Wallach



don't publish online in the dark...



let our open linking network
shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-

profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at www.crossref.org, and make sure all your electronically published work is CrossRef-enabled.



CROSSREF.ORG | THE CITATION LINKING BACKBONE

200 WHEELER ROAD, 6TH FLOOR, BURLINGTON, MA 01803 • 781-359-2442

THEME INTRODUCTION

Advanced Lung Disease: Comorbidities and Challenges

Maria L. Padilla, MD,¹ and Virginia R. Litle, MD²

¹Division of Pulmonary, Critical Care and Sleep Medicine, Mount Sinai School of Medicine, New York, NY

²Division of Thoracic and Foregut Surgery, University of Rochester Medical Center, Rochester, NY

In September 2007, we held a symposium on advanced lung disease at the Mount Sinai Hospital. We focused on the disease entities that result in life-threatening pulmonary dysfunction. We asked our symposium speakers to emphasize the fibrosing lung disorders and to delineate the natural history of these diseases, the radiological features of these entities, and, most importantly, the comorbidities that arise in these diseases, which have a great impact on the quality of life of the patients. At our symposium, the Mount Sinai faculty members and invited guests described the challenges confronted by pulmonologists and surgeons caring for these patients. We learned of potential solutions to such challenges gleaned from the literature and personal clinical experience.

Adam Morgenthau and Maria Padilla described the spectrum of diffuse fibrosing parenchymal lung disease, whereas Alvin Teirstein and Adam Morgenthau delineated the clinical, radiologic, and pathologic changes in advanced sarcoidosis, emphasizing pulmonary fibrosis as an outcome. Virginia Litle, Barry Jaffin, and Yevgenia Pashinsky reviewed gastroesophageal reflux disease as an exacerbating factor in advanced lung disease and detailed the diagnostic tests for gastroesophageal reflux disease and the potential role for antireflux surgery. Joel Nations

and Steven Nathan reviewed the comorbidities of advanced lung disease and discussed current and potential therapeutic interventions. The issues surrounding quality of life and palliative care for patients with advanced lung disease were discussed by Christopher Smith and Cecilia Smith. Pamela Argiriadi and David Mendelson described the radiologic features of the idiopathic interstitial pneumonias and provided a myriad of chest computed tomography images for these uncommon but morbid diseases.

Most importantly, our symposium brought together leaders in this field to exchange thoughts and impressions about idiopathic pulmonary fibrosis. This meeting will be the catalyst for additional clinical research in these life-threatening and debilitating disorders, for which our current treatments are clearly inadequate.

We have asked our contributors to summarize the points that they made in their oral presentations and to prepare articles for publication to allow much wider dissemination of the knowledge that we shared. We hope that you can use the information provided in this issue of the *Mount Sinai Journal of Medicine* to increase your understanding of the clinical management of this challenging patient population.

Address Correspondence to:

Maria L. Padilla

Division of Pulmonary, Cardiac
Care and Sleep Medicine Mount
Sinai School of Medicine
New York, NY
Email: maria.padilla@
mountsinai.org

DISCLOSURES

Potential conflict of interest: Nothing to report.

Spectrum of Fibrosing Diffuse Parenchymal Lung Disease

Adam S. Morgenthau, MD, FCCP, and Maria L. Padilla, MD, FCCP

Division of Pulmonary, Critical Care, and Sleep Medicine, Mount Sinai School of Medicine, New York, NY

ABSTRACT

The interstitial lung diseases are a heterogeneous group of disorders characterized by inflammation and/or fibrosis of the pulmonary interstitium. In 2002, the American Thoracic Society and the European Respiratory Society revised the classification of interstitial lung diseases and introduced the term *diffuse parenchymal lung disease*. The idiopathic interstitial pneumonias are a subtype of diffuse parenchymal lung disease. The idiopathic interstitial pneumonias are subdivided into usual interstitial pneumonia (with its clinical counterpart idiopathic interstitial pneumonia), nonspecific interstitial pneumonia, cryptogenic organizing pneumonia, acute interstitial pneumonia, desquamative interstitial pneumonia, respiratory bronchiolitis interstitial lung disease, and lymphocytic pneumonia. Sarcoidosis and hypersensitivity pneumonitis are the 2 most common granulomatous diffuse parenchymal lung diseases. Rheumatoid arthritis, systemic sclerosis, and dermatomyositis/polymyositis (causing antisynthetase syndrome) are diffuse parenchymal lung diseases of known association because these conditions are associated with connective tissue disease. Hermansky-Pudlak syndrome is a rare genetic diffuse parenchymal lung disease characterized by the clinical triad of pulmonary disease, oculocutaneous albinism, and bleeding diathesis. This review provides an overview of the chronic fibrosing diffuse parenchymal lung diseases. Its primary objective is to illuminate the clinical challenges encountered by clinicians who manage the diffuse parenchymal lung diseases regularly and to offer potential solutions to those challenges. Treatment for the diffuse parenchymal

lung diseases is limited, and for many patients with end-stage disease, lung transplantation remains the best option. Although much has been learned about the diffuse parenchymal lung diseases during the past decade, research in these diseases is urgently needed. *Mt Sinai J Med* 76:2–23, 2009. © 2009 Mount Sinai School of Medicine

Key Words: antisynthetase syndrome, dermatomyositis, diffuse parenchymal lung disease, Hermansky-Pudlak syndrome, hypersensitivity pneumonitis, idiopathic pulmonary fibrosis, nonspecific interstitial pneumonia, polymyositis, sarcoidosis, usual interstitial pneumonia.

The interstitial lung diseases (ILDs) are a heterogeneous group of disorders characterized by inflammation and/or fibrosis of the pulmonary interstitium. Progression of the diseases results in impaired oxygen transfer and scarring within the lung. The ILDs were originally classified by their histologic patterns into clinically relevant groups to reflect that each disorder could occur in a variety of clinical contexts.¹ In 2002, the American Thoracic Society (ATS) and European Respiratory Society (ERS) revised the original classification scheme² and introduced the term *diffuse parenchymal lung disease* (DPLD) in place of the term *ILD* to emphasize that the parenchyma is the primary site of injury in DPLDs and to distinguish the DPLDs from pulmonary conditions predominantly affecting the airways and the blood vessels.

The guidelines (Figure 1) classify these disorders into 4 categories: (1) DPLDs of known association (eg, drugs, connective tissue diseases, and environmental/occupational exposures), (2) granulomatous DPLDs (eg, sarcoidosis), (3) other and often rare DPLDs (eg, lymphangioleiomyomatosis, pulmonary Langerhans cell histiocytosis, pulmonary alveolar proteinosis, and eosinophilic pneumonia), and (4) the idiopathic interstitial pneumonias (IIPs).² The IIPs are further subdivided into usual interstitial pneumonia [UIP; with its clinical counterpart idiopathic pulmonary fibrosis (IPF) when no known cause is identified for the UIP pattern], nonspecific interstitial

Address Correspondence to:

Adam S. Morgenthau
Division of Pulmonary, Critical
Care, and Sleep Medicine
Mount Sinai School of Medicine
New York, NY
Email: adam.morgenthau@mssm.edu

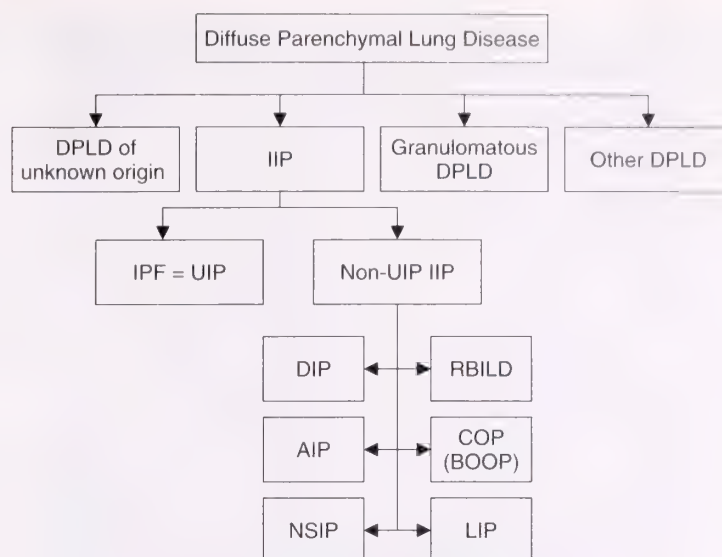


Fig 1. DPLDs consist of disorders of known causes (eg, collagen vascular disease or environmental or drug-related causes) and disorders of unknown causes. The latter include IIPs, granulomatous lung disorders, and other forms of interstitial lung disease. The most important distinction among the IIPs is that between IPF and NSIP. Other IIPs include DIP, RBILD, AIP, COP, and LIP. **Abbreviations:** AIP, acute interstitial pneumonia; COP, formerly known as Bronchiolitis obliterans organizing pneumonia (BOOP); COP, cryptogenic organizing pneumonia; DIP, desquamative interstitial pneumonia; DPLD, diffuse parenchymal lung disease; IIP, idiopathic interstitial pneumonia; IPF, idiopathic pulmonary fibrosis; LIP, lymphocytic interstitial pneumonia; NSIP, nonspecific interstitial pneumonia; RBILD, respiratory bronchiolitis–associated interstitial lung disease; UIP, usual interstitial pneumonia. The ATS/ERS Clarification of Diffuse Parenchymal Lung Disease.

pneumonia (NSIP), cryptogenic organizing pneumonia, acute interstitial pneumonia, desquamative interstitial pneumonia, respiratory bronchiolitis ILD, and lymphocytic interstitial pneumonia.

This review provides an overview of the fibrosing DPLDs, examining several of the disorders within each of the 4 DPLD categories. We discuss fibrotic NSIP and idiopathic UIP/IPF. We review the 2 most common granulomatous DPLDs, sarcoidosis and hypersensitivity pneumonitis (HP; extrinsic allergic alveolitis). We examine 3 connective tissue disorders [rheumatoid arthritis (RA), systemic sclerosis (SSc), and dermatomyositis (DM)/polymyositis (PM) causing antisynthetase syndrome (AS)] with frequent parenchymal lung involvement. We also discuss a rare genetic DPLD known as Hermansky-Pudlak syndrome (HPS).

The review illuminates the clinical challenges encountered by clinicians who regularly manage end-stage DPLDs and offers potential solutions to some of these challenges. For many patients, lung transplantation remains the ultimate treatment option. However, only 40% of lung transplantation patients survive

5 years from the date of their transplant.³ Although much has been learned about the DPLDs in the recent past, research on these diseases is urgently needed

IDIOPATHIC INTERSTITIAL PNEUMONIAS

UIP is a distinct histopathologic pattern observed in IPF and several other DPLDs.⁴ It is typically diagnosed by surgical lung biopsy, but a high-resolution computed tomography (HRCT) scan may be diagnostic if the radiologic features of UIP are classic for the disease.⁵ The classification system devised by the ATS/ERS in 2002 (Figure 1) emphasizes the importance of an integrated clinical, radiologic, and histopathologic approach to the diagnosis of IIPs.²

Idiopathic Pulmonary Fibrosis/Usual Interstitial Pneumonia

IPF/UIP is a disease of uncertain etiology, characterized by progressive dyspnea on exertion, “velcro”

Table 1. *Criteria for Diagnosis of Idiopathic Pulmonary Fibrosis in the Absence of Surgical Lung Biopsy.*

Major criteria (all 4 features must be present)

- Exclusion of other known causes of interstitial lung disease, such as drug toxicities, environmental exposures, and collagen vascular diseases
- Abnormal pulmonary function studies that include evidence of restriction with or without impaired gas exchange
- Bibasal reticular abnormalities with minimal ground glass opacities on a high-resolution computed tomography scan
- Transbronchial lung biopsy or bronchoalveolar lavage showing no features to support an alternative diagnosis

Minor criteria (3 of 4 features must be present)

- Age > 50 years
- Insidious onset of otherwise unexplained dyspnea on exertion
- Duration of illness: 3 months
- Bibasal inspiratory crackles (dry or "velcro" in quality)

lung crackles on chest auscultation, and escalating hypoxemia with eventual oxygen dependence. Approximately half of all patients have clubbing of the fingertips. The major and minor diagnostic criteria are shown in Table 1. Men are affected more commonly than women. Rare familial cases of IPF have been described. Survival in UIP/IPF is poor. Greater than 50% of patients with the disease die 2 to 3 years after their diagnosis.⁶ In almost all cases, there is a progressive disease course punctuated by worsening symptoms with an accelerated decline on pulmonary function tests.⁷

The classic HRCT scan (Figure 2) in IPF/UIP reveals coarse reticular or linear opacities in the basilar and/or peripheral (subpleural) regions of the lungs. Subpleural honeycomb cysts are also typically present. In some cases, there is thickening of the intralobular/interlobular septae as well as traction bronchiectasis. Diffuse ground-glass opacification (GGO) of the parenchyma is rare, but focal GGO is fairly common.

Temporal heterogeneity is the central histopathologic feature that characterizes UIP and distinguishes it from other IIPs, especially NSIP.^{8,9} The term *temporal* refers to lung injury that occurs at various points in time in a chronological fashion. The UIP pattern (Figure 3) is heterogeneous because lung injury appears in different pathological stages.^{4,8,9} Other salient histological features of UIP/IPF include honeycombing, sparse cellular inflammation, and areas of proliferating myofibroblasts known as fibroblastic foci.⁴

Fibrotic Nonspecific Interstitial Pneumonia

Dyspnea and cough, the most common symptoms of NSIP, generally develop subacutely.^{7,10} About one-third of patients have fever and/or constitutional

symptoms. Patients report dry mouth or eyes, joint or muscle pain, muscle weakness, Raynaud's phenomenon, or rash if an underlying collagen vascular disease is present. Most patients have bibasilar crackles, but only 10% to 35% have clubbing.¹¹ Pulmonary function testing typically demonstrates a restrictive ventilatory defect and a decrease in gas transfer.²

Fibrotic NSIP is 1 of 2 histopathologic patterns of NSIP (the other pattern, cellular NSIP, is not discussed in this review), which may be difficult to distinguish from UIP/IPF but is paradoxically defined by its dissimilarity to UIP/IPF (Table 2). Fibrotic NSIP affects young females¹² and is more common in patients with connective tissue disease.¹³ Patients with fibrotic NSIP almost always have a better prognosis than those with UIP/IPF.¹⁴

HRCT in NSIP (Figure 4) commonly demonstrates GGO that is bilateral and symmetrical.^{5,14,15} It may be associated with volume loss (especially in the lower lobes) and/or traction bronchiectasis. Honeycomb change is rare.^{7,10,16}

The histopathology in fibrotic NSIP (Figure 5) is temporally homogeneous.¹⁶ There is uniform distribution of inflammation and/or fibrosis within the interstitium, and lung injury occurs at the same pathological stage. Fibroblastic foci and honeycomb change, salient histopathologic features of UIP, are usually absent in fibrotic NSIP. In cases in which the histopathology does not differentiate NSIP from UIP, clinical and imaging features determine diagnosis.

Is High-Resolution Computed Tomography Helpful in Differentiating Usual Interstitial Pneumonia from Fibrotic Nonspecific Interstitial Pneumonia?

HRCT has dramatically altered the diagnostic evaluation of patients with DPLD. However, can it be used

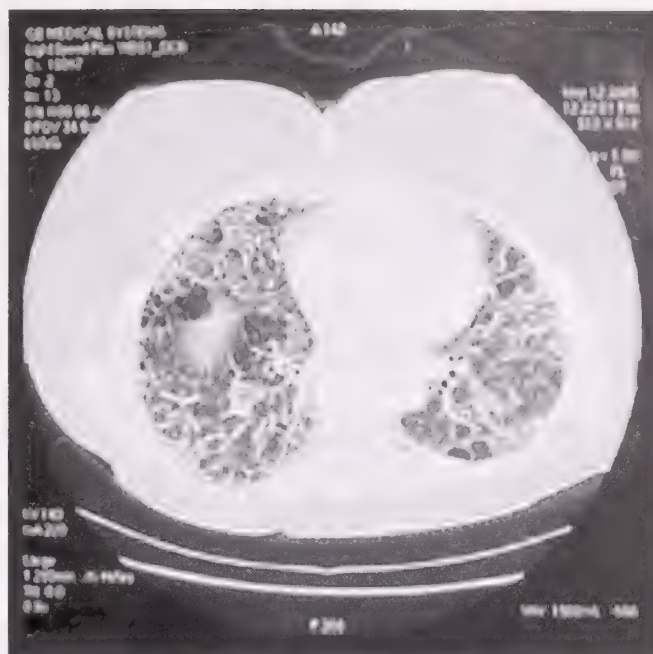


Fig 2. High-resolution computed tomography scan (from the same patient) showing characteristic interlobular and intralobular septal thickening, ground-glass attenuation, loss of volume, traction bronchiectasis, and honeycombing with a subpleural bibasal predominance.

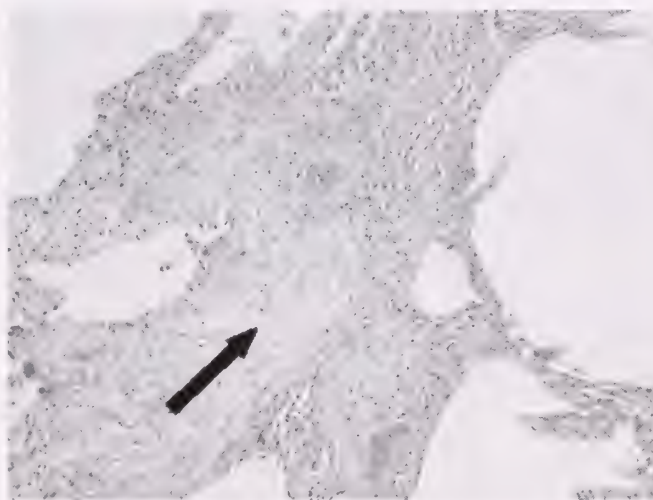


Fig 3. Typical histological features of usual interstitial pneumonia. The patchy nature of usual interstitial pneumonia is clearly demonstrated. There are relatively normal alveoli intermixed with significant interstitial inflammation. A fibroblast focus (arrow) can also be visualized. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

to discriminate fibrotic NSIP from UIP/IPF? Hunninghake and colleagues¹⁷ examined the diagnostic value of HRCT in the clinical, radiographic, and pathologic diagnosis of 91 patients suspected of having IPF. The histopathology served as the gold standard for diagnosis. When a group of radiologists and a separate

group of clinicians evaluated HRCT, the positive predictive value of a UIP diagnosis in each group was 85% and 87%, respectively. When the radiologists felt confident about their HRCT diagnosis (approximately 60% of the cases that they evaluated), the positive predictive value improved to 96%.⁵ Subpleural

Table 2. *Clinical Features Differentiating IPF from NSIP.*

Clinical/radiographic/pathological features	IPF	NSIP
Duration of illness	Chronic (>12 months)	Subacute to chronic (months to years)
Frequency of diagnosis	47–64%	14–36%
Chest radiograph	Bilateral reticular opacities in lower zones; volume loss; $\pm$ honeycombing	Bilateral hazy and reticular opacity
HRCT	Peripheral, subdeural, basal predominance Reticular opacities Honeycombing Traction bronchiectasis Architectural distortion. Focal ground-glass	Peripheral, subpleural, basal, symmetrical Ground-glass attenuation Consolidation Lower lobe volume loss
Key histological features	Usual interstitial pneumonia pattern Dense fibrosis causing remodeling of lung architecture with frequent “honeycomb” fibrosis Fibroblastic foci typically scattered at the edges of dense scars Patchy lung involvement Frequent subpleural and paraseptal distribution	NSIP pattern Cellular pattern Mild to moderate interstitial chronic inflammation Type II pneumocyte hyperplasia in areas of inflammation Fibrosing pattern Dense or loose interstitial fibrosis lacking the temporal heterogeneity pattern and/or patchy features of UIP Lung architecture may appear lost on examination of H&E stained sections but relatively preserved with elastic stains Interstitial chronic inflammation (mild or moderate)
Treatment	Poor response to any treatment	Corticosteroid responsiveness
Prognosis	50–70% mortality in 5 years	Unclear; <15% mortality in 5 years

NOTE: This table has been reprinted with permission from *Thorax*.¹⁷⁸ Copyright 2007, British Thoracic Society. The caption has been slightly modified, and the abbreviation footnote has been added. **Abbreviations:** HRCT, high-resolution computed tomography; IPF, idiopathic pulmonary fibrosis; NSIP, nonspecific interstitial pneumonia.

honeycombing and septal thickening were the HRCT features most predictive of IPF. Patients with HRCT scans that exhibited these features suffered the worst survival.¹⁵

In contrast to IPF/UIP, it is difficult to diagnose NSIP solely on the basis of its radiologic features,^{5,15,18,19} and a radiologic diagnosis of NSIP must be made with caution. Ground-glass attenuation is the only HRCT feature independently associated with NSIP.^{5,15,18,19} Most experts agree that in patients who are good surgical candidates, surgical lung biopsy should be performed to confirm the diagnosis of NSIP.

Is It Prognostically Significant When Multiple Biopsy Specimens Exhibit Discordant Histopathology and/or When the Histopathology Is Discordant with Radiologic Features on High-Resolution Computed Tomography?

To reduce the sampling error and increase the specificity of open lung biopsy, multiple specimens should be obtained from different sites within the lung.^{6,9,20} In some cases, the histopathology of these specimens is discordant. Several studies have examined the impact that discordant histopathology has on prognosis. Flaherty and associates¹⁵ demonstrated

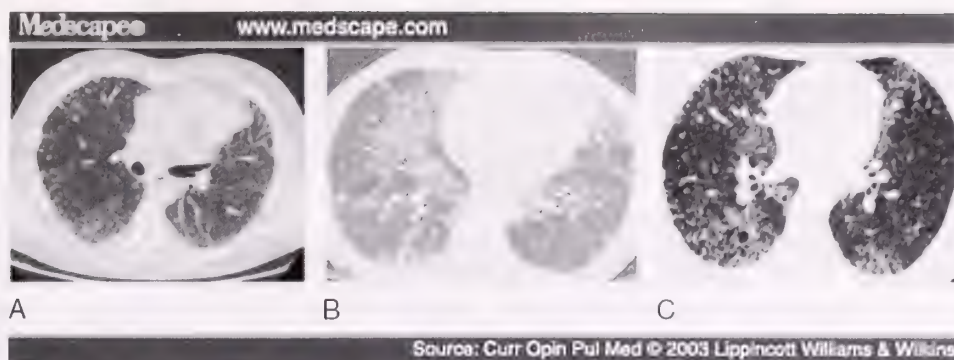


Fig 4. High-resolution computed tomography scans from a patient with nonspecific interstitial pneumonia. There is less peripheral distribution of lesions in comparison with usual interstitial pneumonia/idiopathic pulmonary fibrosis. There is also widespread ground-glass change as well as traction bronchiectasis, but there is little honeycombing. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]



Fig 5. Low-magnification photomicrograph showing nonspecific interstitial pneumonia in which a combination of inflammation and fibrosis expands alveolar septa in a uniform fashion without associated architectural distortion. Reprinted with permission from *Proceedings of the American Thoracic Society*.¹⁷⁷ Copyright 2006, Proceedings of the American Thoracic Society. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

that patients with discordant UIP-NSIP histopathology had a poor prognosis that was similar to the prognosis of those with concordant UIP. Patients with concordant NSIP histopathology had a significantly better outcome than those with either discordant UIP-NSIP or concordant UIP. Interestingly, concordant NSIP was more prevalent in younger patients, and patients with discordant UIP-NSIP were younger than those with concordant UIP. The reason for the age differences is not understood.

However, what is the prognostic significance of discordant radiologic (HRCT) and histopathologic

findings? In 1 study, NSIP patients with concordant studies demonstrated a better prognosis than UIP patients with concordant studies.¹⁷ Patients with discordant studies (HRCT consistent with NSIP and biopsy demonstrating UIP) exhibited an intermediate survival that was higher than that in patients with concordant UIP but lower than that in patients with concordant NSIP. Above all, the study emphasized the importance of identifying the cohort of UIP patients with concordant histopathology and computed tomography (CT). This group of patients is least likely to benefit from treatment and, therefore, has the worse prognosis.

Are There Biological Markers of Disease Activity in Idiopathic Pulmonary Fibrosis/Usual Interstitial Pneumonia and Nonspecific Interstitial Pneumonia?

Numerous studies have attempted to identify potentially relevant biological markers of disease activity in IPF/UIP and/or NSIP.^{21–32} KL-6 is a high-molecular-weight glycoprotein that is elevated in patients with several types of ILDs and may have prognostic significance in these disorders.^{21,31} Plasma D-dimer and anti-endothelial antibodies are 2 promising candidate biomarkers.

Plasma D-dimer levels have been associated with mortality in IPF.³³ Kubo³³ indirectly evaluated the effect of anticoagulant therapy on the survival of patients with IPF. Patients who received either prednisolone alone or prednisolone and anticoagulant therapy were monitored for acute exacerbations of their disease. The mortality associated with an acute exacerbation of IPF was reduced in patients who received anticoagulation ($P = 0.008$ and $P < 0.0001$, respectively). The author concluded that anticoagulant therapy improved survival in patients with IPF. Additional research is needed to validate the results of the study.

Anti-endothelial cell antibodies (AECAs) are a heterogeneous group of antibodies that recognize various antigenic determinants on human endothelial cells. AECAs are present in patients with collagen vascular diseases (CVDs) and are associated with CVD-related ILDs.^{2,6,9,16} Matsui and colleagues³⁴ investigated the prevalence of AECAs in patients with lung fibrosis and examined whether the expression of these antibodies differed among the IIPs: UIP, NSIP, and CVD-associated ILD. Sera from patients with IPF/UIP did not contain AECAs, but the antibodies were present in 5 of 10 patients with idiopathic NSIP, 5 of 14 patients with CVD-UIP, and 4 of 10 patients with CVD-NSIP ($P < 0.05$). The authors concluded that AECAs have a pathogenic role in the fibrotic process and that each IIP might have a unique pathogenesis.

What Are the Prognostic Factors Associated with Idiopathic Pulmonary Fibrosis/Usual Interstitial Pneumonia and/or Fibrotic Nonspecific Interstitial Pneumonia?

The factors that govern prognosis in IPF/UIP and/or fibrotic NSIP can be divided into 4 categories: (1) clinical prognostic factors, (2) physiologic prognostic factors, (3) radiologic prognostic factors, and (4) histopathologic factors. The radiologic factors

implicated in the prognosis of these disorders has been discussed previously.

The clinical prognostic factors associated with IPF/UIP concern the comorbid conditions associated with the disease. Pulmonary hypertension complicates IPF/UIP, and the prognosis is adversely affected.^{35–37} The size of the PA segment, as measured on the plain chest radiograph, correlates with mortality.³⁸ The systolic pulmonary artery pressure (sPAP), estimated by echocardiography, strongly correlates with survival.³⁵ Patients with an estimated sPAP greater than 50 mm Hg had a median survival of only 0.7 years versus more than 4 years for those patients with an sPAP of less than 50 mm Hg. In a more recent study, 79 IPF patients underwent right-heart catheterization. There was a significant difference in the 1-year mortality rate in those patients with PH (28%) versus those without it (5.5%; $P = 0.002$).³⁷

Acute exacerbations in IPF adversely affect prognosis in IPF/UIP.^{39–47} Kim and colleagues⁴³ retrospectively evaluated 147 patients with IPF, 11 of whom met criteria for an acute exacerbation. These included (1) aggravation of dyspnea within 1 month, (2) hypoxemia with an arterial oxygen tension/inspired oxygen tension ratio of <225 , (3) newly developing pulmonary infiltrates on chest radiography, and (4) absence of apparent infection or heart disease. The 2-year incidence of an acute exacerbation was 9.6%, which was associated with a mortality rate of 78%. The acute exacerbations occurred between 3 and 60 months after the time of the initial visit.

A randomized, double-blinded, placebo-controlled trial of pirfenidone for the treatment of IPF evaluated the frequency of acute exacerbations as a secondary endpoint.⁴⁸ Patients were observed for 6 months. Five of 107 patients developed exacerbations, and all of them were members of the placebo group. The untreated IPF patients appeared to have an increased incidence of acute exacerbations. However, the study was discontinued after 6 months because there was an increased incidence of exacerbations in 1 of the treatment cohorts.

Progressive dyspnea affects prognosis in IPF/UIP.^{49–53} In 1 study, patients' dyspnea on exertion was scored on a scale of 0 to 20.⁵⁴ The baseline dyspnea score predicted survival: each 1-point change was associated with an 11% difference in survival time after adjustments for age and smoking status ($P = 0.008$).

Physiologic factors have an impact on prognosis in IPF/UIP. Collard and colleagues⁵⁴ showed that the baseline total lung capacity (TLC), thoracic gas volume, forced vital capacity (FVC), forced expiratory

volume in 1 second (FEV1), and diffusing capacity of the lung for carbon monoxide (DL_{CO}) predicts survival time. In the same study, the baseline partial pressure of arterial oxygen (PaO₂), oxygen saturation, and alveolar-arterial oxygen gradient also predicted survival time.⁵⁴

A serial decline in FVC accurately predicted mortality in IPF patients who did not desaturate to less than 88% during 6-minute-walk testing.⁵⁵ The DL_{CO} parameter, in contrast, more consistently predicted mortality in patients who desaturated to less than 88% during the test.

The 6-minute-walk test may risk-stratify and prioritize patients for lung transplantation.⁵⁶ Transplant candidates with IPF who walked a distance of less than 207 m exhibited a 4-fold increase in mortality.

Histopathologic factors affect prognosis in IPF/UIP. A high profusion of fibroblastic foci in UIP is associated with higher mortality in patients with IPF.^{55,57,58} Recent studies have suggested that fibroblastic foci embody a complex reticulum that extends from the pleura into the underlying parenchyma.⁵⁹ Conceivably, the fibroblastic reticulum impairs gas transfer, and this adversely affects mortality. However, further studies are needed to elucidate the impact of histopathologic factors on prognosis in the absence of abundant fibroblastic foci.

How Should Patients with Idiopathic Pulmonary Fibrosis/Usual Interstitial Pneumonia or Fibrotic Nonspecific Interstitial Pneumonia Be Treated?

In 2002, an ATS/ERS international consensus statement concluded that "no data exist that adequately document any current treatment improves survival or the quality of life for patients with IPF."² In our view, it is clinically useful to divide the treatments for IPF/UIP into 4 categories: (1) conventional treatment, (2) no treatment, (3) experimental treatment within a clinical trial, and (4) lung transplantation.

There are remarkably few data that support the use of immunosuppressive agents for IPF/UIP.^{16,60} Oral corticosteroids are efficacious in some patients.¹⁴ Flaherty *et al.*¹⁴ treated 29 IPF/UIP patients with oral corticosteroid therapy and measured the response to therapy using a combined clinical, pathologic, and physiologic (C-reactive protein) score. Seventeen percent of the patients responded to treatment, 31% remained stable, and 52% did not respond. All treated patients experienced a serious adverse event. The therapy was not associated with a survival benefit. Randomized, placebo-controlled trials are required to determine the true effects of corticosteroid therapy in IPF.

Anti-inflammatory medicines have been traditionally used to treat NSIP. However, the impact of therapy is difficult to ascertain because prospective, randomized, controlled trials do not exist. The majority of NSIP patients treated with anti-inflammatory agents improve or remain stable on therapy, but relapse is common. Cellular NSIP has responded more favorably to anti-inflammatory medicines than fibrotic NSIP. Martinez and colleagues¹⁴ evaluated 10 NSIP patients and 29 IPF patients treated with high-dose prednisone (1 mg/kg/day for 1–3 months with a subsequent taper). According to the clinical Radiographic Physiologic score 4 of 10 (40%) patients with NSIP improved, 5 patients remained stable, and 1 patient deteriorated.

Conventional treatments for IPF/UIP and fibrotic NSIP have limited efficacy. Several novel potential therapies have been identified to treat these disorders (Table 3).⁶¹ It is beyond the scope of this review to discuss all of these agents. However, it is important to consider several therapies that are being evaluated in current clinical trials because many patients with IPF/UIP or fibrotic NSIP fail conventional therapy or are inappropriate candidates for the therapy. In our view, most of these patients should be enrolled into clinical trials.

The efficacy of pirfenidone (an oral antifibrotic agent) was evaluated in a double-blind, placebo-controlled, multicenter study of 107 patients with IPF.⁴⁸ Six months after initiation of pirfenidone, none of the patients experienced significant improvement in their oxygen saturation on 6-minute walk testing. Nine months after commencing the drug, patients with milder disease had an improvement (that was not statistically significant) in the vital capacity and the frequency of acute exacerbations. Pirfenidone is currently being evaluated in phase II and III clinical trials.

Interferon- γ (IFN- γ) is a cytokine with antifibrotic, anti-inflammatory, and anti-infective effects.^{60,62} It was first evaluated for the treatment of IPF in a pilot study.⁶³ Patients treated with IFN- γ -1b demonstrated improved total lung capacity and increased PaO₂ at rest ($P < 0.001$).⁶⁴

In a double-blind, placebo-controlled trial conducted at 58 centers in the United States, Europe, Canada, and South Africa, 330 patients were assigned randomly to treatment with IFN- γ -1b (200 μ g subcutaneously 3 times per week) or placebo for 48 weeks.⁶⁵ Eligible patients exhibited the following clinical symptoms for at least 3 months, an FVC that was 50% to 90% of the predicted value, a DL_{CO} $\geq$ 25% of the predicted value, and a PaO₂ $>$ 55 mm Hg at rest. Participants also had to demonstrate deterioration during the preceding

Table 3. *Candidate Therapeutic Approaches According to Targeted Pathways in Idiopathic Pulmonary Fibrosis.*

Alveolar epithelial cells	• Apoptosis (ACE-i, angiotensin type I receptor antagonists) • Epithelial replacement (KGF, HGF) • Procoagulant activity (PAI) • GF production/inhibition ◦ Anti-TGF-β therapy: anti-TGF Abs (CAT-192, CAT-152, GC1008), TGF-β receptor kinase inhibitors (PKI), inhibitors of post-receptors protein kinases, decorin ◦ Etanercept: phase II (completed) ◦ PAF-receptor antagonists (WEB 2086), anti-integrin Abs ◦ Imatinib mesylate: phase II (ongoing) ◦ FG-1019: phase I (completed), phase II (ongoing) ◦ Suramin
Fibroblasts myofibroblasts	• Matrix turnover (pirfenidone, relaxin) • Inhibitors of fibroblast migration/proliferation ◦ Zileuton: (phase II, ongoing) ◦ PGE₂, PGI₂ analog, sildenafil ◦ Pirfenidone [phase II (completed), phase III (planned)] • Angiotensin II inhibitors • Inductors of fibroblast/myofibroblasts apoptosis (statins) • Deactivation of myofibroblasts (lufironil, safronil)
Angiogenesis	• Bosentan • Angiostatic chemokines • IFN-γ-1b: phase III (completed), second phase III (recruiting) ◦ Tetrathiomolybdate: phase I/II (completed)
Th1/Th2-type cytokine balance	• IFN-γ-1b • anti-IL-4 • anti-IL-13
Oxidative stress	• Cysteine pro-drugs [NAC: phase III (completed), Z2196] • Pirfenidone • Tetrathiomolybdate: phase I/II (completed) • Taurine and niacin • NF-κB antisense oligonucleotides • Nitric oxide synthase inhibitors (aminoguanidine)

NOTE: This table has been reprinted with permission from *Pulmonary Pharmacology & Therapeutics*.⁶¹ Copyright 2007, Elsevier, Ltd. The caption and abbreviation footnote have been slightly modified. **Abbreviations:** Ab, antibody; ACE-i, angiotensin-converting enzyme inhibitors; GF, growth factor; HGF, hepatocyte growth factor; KGF, keratinocyte growth factor; IFN- γ , interferon- γ ; IL, interleukin; NAC, *N*-acetylcysteine; NF- κ B, nuclear factor- κ B; PAF, plasminogen activating factor; PAI, plasminogen activator inhibitor; PGE₂, prostaglandin E₂; PGI₂, prostaglandin I₂; PKI, protein kinase inhibitor; TGF- β , transforming growth factor- β ; Th1, T helper 1; Th2, T helper 2.

year (a decrease in predicted FVC $\geq$ 10%, worsening features on chest radiograph, or worsening dyspnea) despite steroid therapy. Three criteria defined the primary endpoint: (1) a decline in FVC $>$ 10%, (2) an increase in the alveolar-arterial oxygen tension difference $>$ 5 mm Hg, and (3) death.

With respect to the placebo, IFN γ -1b did not significantly affect progression-free survival. Disease progression or death occurred in 46% of patients in the IFN γ -1b group and in 52% of patients in the placebo group.⁶³ There was a possible mortality benefit in treated patients with milder disease. Subsequent studies did not confirm the positive trend.

Oxidant/antioxidant imbalance is central to the pathogenesis of IPF.⁶⁵ Glutathione functions as an antioxidant, and reduced levels are found in the lungs

of patients with IPF.⁶⁶ *N*-Acetylcysteine is a precursor for glutathione synthesis. Its therapeutic role in IPF was evaluated in a double-blind, randomized, placebo-controlled study. When it was administered with prednisolone and azathioprine, it preserved vital capacity and gas to a greater extent than standard therapy alone (relative difference of 9% for FVC and 24% for gas transfer). However, FVC and DL_{CO} declined in both groups.⁶⁷ The difference in the extent of the decline reached statistical significance, but there was not a significant mortality benefit in the group treated with *N*-acetylcysteine.

The endothelins are a family of 21-amino-acid peptides that have profibrotic properties. Endothelin-1 is expressed in a variety of pulmonary diseases, including pulmonary vascular disease and pulmonary

fibrosis.⁶³ More recently, it was evaluated in the treatment of IPF [the BUILD 1 trial (bosentan use in ILD) and the BUILD 2 trial (bosentan use in ILD associated with systemic disorders)].⁶⁵ In the BUILD 1 trial, bosentan reduced the incidence of death or treatment failure at 12 months; this was a positive effect that was not statistically significant. Results of the BUILD 2 trial are forthcoming. Participants are currently being recruited for the BUILD 3 trial. The primary goal of the study is to determine whether bosentan delays disease progression or death in patients with IPF.

Tumor necrosis factor- α (TNF- α) is a proinflammatory cytokine involved in the pathogenesis of pulmonary fibrosis.⁶⁷ It has a variety of profibrotic effects, such as stimulation of fibroblast proliferation and collagen gene up-regulation.⁶⁸ Mice missing TNF- α receptors or treated with soluble TNF- α receptors are relatively resistant to bleomycin/silica-induced fibrosis.^{68,69} In humans, TNF- α gene polymorphisms have been localized to chromosome 6. One of these polymorphisms has been associated with an increased risk of developing IPF.⁷⁰

Etanercept is a recombinant soluble TNF- α receptor blocker. One study prospectively assessed the effects of etanercept in 9 patients with advanced IPF (the mean FVC was 46% of that predicted). After an average of 9 months of treatment (etanercept and 10 mg of prednisolone twice weekly), 1 patient died, 3 demonstrated a 15% increase in FVC from baseline, and 4 showed an improvement in gas transfer. The results of the recently completed phase II, multicenter, randomized, double-blind, placebo-controlled study are eagerly anticipated.

Imatinib mesylate (Gleevec), a potent and specific tyrosine kinase inhibitor that is used in patients with chronic myeloid leukemia and gastrointestinal tumors, is one of the key growth factors that plays a role in the pathogenesis of pulmonary fibrosis.⁷¹ Imatinib has been shown to inhibit proliferation of mesenchymal cells in a model of bleomycin-induced pulmonary fibrosis.^{71,72} A double-blind, placebo-controlled phase II-III trial of imatinib in patients with IPF has been completed, but the results of the study have not yet been published.⁷²

Novel therapies abound, but treatment for UIP/IPF should be individualized, and a frank discussion of the risks and benefits associated with therapy is essential. No agent has been shown to influence survival, and each of these medicines has adverse effects. Supportive therapies for IPF are important and are often overlooked. Pulmonary rehabilitation and the treatment of comorbid illnesses may ameliorate patients' quality of life and are usually the most appropriate treatments for patients with UIP/IPF.⁷² No treatment may be the best treatment

for certain patients such as elderly patients who exhibit classic features of UIP on a CT scan and a gradually deteriorating course.

Lung transplantation is the only therapeutic intervention that offers patients with IPF a survival benefit.^{73,74} Pulmonary fibrosis represents the second most frequent disease for which lung transplantation is performed, but IPF patients have the poorest survival on the transplantation waiting list.^{71-73,75} Several studies suggest that patients with pulmonary fibrosis have better outcomes than those who undergo transplantation for other pulmonary conditions.^{76,77} The optimal time for listing these patients for transplantation is uncertain. However, referral and evaluation at the time of diagnosis increase the likelihood that patients who develop an acute exacerbation will undergo transplantation in a timely manner.

GRANULOMATOUS DIFFUSE PARENCHYMAL LUNG DISEASES

Sarcoidosis is a multisystem disease of uncertain etiology that is characterized by several associations with specific gene products and the presence of noncaseating epithelioid granulomas. Granuloma formation is a pathologic response to primarily foreign material. Granulomas form around a nidus of poorly soluble or insoluble material that cannot be simply removed by a single cell. They confine pathogens, restrict inflammation, and protect surrounding tissue. In sarcoidosis, granuloma formation disrupts tissue homeostasis and impairs organ function.

The lungs are the most frequent site of organ involvement in sarcoidosis.⁷⁸⁻⁸¹ Abnormalities on chest radiographs are detected in 85% to 95% of patients with the disease.⁸¹⁻⁸³ Patients with pulmonary sarcoidosis commonly have dyspnea and cough. This review on sarcoidosis complements the discussion of fibrotic, end-stage sarcoidosis found in this issue of the *Mount Sinai Journal of Medicine*.

Diagnosis

The diagnosis of sarcoidosis is established by clinical and radiographic findings that are compatible with the disease and are supported by histologic evidence of noncaseating epithelioid granulomas (which do not stain positively for microorganisms or other particles) in 1 or more organs. Most experts agree that patients with Löfgren's syndrome and classic radiographic findings do not need to be biopsied.⁸⁴ In all other cases, a biopsy specimen should be

obtained from the involved organ that is most easily accessed. The skin and peripheral lymph nodes are commonly biopsied, but bronchoscopic transbronchial lung biopsy is frequently employed for diagnosis. In experienced hands, the diagnostic yield of transbronchial lung biopsy is greater than 85% when multiple lung segments are sampled.⁸⁴

Sarcoidosis can also be diagnosed by the Kveim-Siltzbach test. The test is performed by the intradermal injection of homogenate of human sarcoid tissue extract. After 4 weeks, the papule that develops at the site of injection is biopsied (Figure 6). The test is not universally available, but it may be used to confirm the diagnosis when it is uncertain or in patients who have lesions that are difficult to biopsy.⁸⁴

Prognosis

The chest radiographic staging system for sarcoidosis has modest prognostic value.⁸⁵ It defines the following stages: stage 0 (normal), stage I [bilateral hilar lymphadenomegaly (BHL) without pulmonary infiltrates], stage II (BHL plus pulmonary infiltrates), stage III (parenchymal infiltrates without BHL), and stage IV (extensive fibrosis with distortion or bullae). The incidence of radiographic stages differs according to geographic regions, ethnicity, and referral bias. Most Scandinavian studies have cited a predominance of radiographic stage I and II disease.^{83,86} Studies from the United States and British Isles report a disproportionate number of cases consistent with radiographic stage III and IV disease.^{87,88}

The prognosis is best with radiographic stage I, intermediate with stage II, and worst with stage

III or IV. Spontaneous remissions occur in 60% to 90% of patients with stage I disease, in 40% to 70% with stage II, in 10% to 20% with stage III, and in 0% with stage IV.^{83,85,86,89,90} Scadding⁸⁵ followed patients with sarcoidosis for 5 years and found that 31 of 32 patients (97%) with stage I disease were asymptomatic. Only 58% of stage II patients and 25% of stage III patients were asymptomatic. Siltzbach⁹⁰ reported similar findings in a population of 244 American patients, some of whom were treated. The chest radiographs normalized in 54% of patients with stage I disease, in 31% with stage II disease, and in 10% with stage III disease. None of 110 patients with a stage I chest radiograph died.

The clinical course and prognosis of sarcoidosis are influenced by ethnic and genetic factors.^{86,87} Black race is associated with a higher rate of chronic progressive disease, worse long-term prognosis, extrapulmonary involvement, and higher risk of relapses.^{80,87,91} ACCESS (A Case Control Etiology Study of Sarcoidosis) demonstrated that women were more likely to have eye and neurological involvement and erythema nodosum. Men were more likely to develop hypercalcemia.⁸¹ There was a higher prevalence of eye, liver, bone marrow, extrathoracic lymph node, and skin involvement (other than erythema nodosum) in black subjects. Derangements in calcium metabolism were more common among white subjects.⁸¹ The influence of human leukocyte antigen (HLA) markers and prognosis is controversial.^{78,92,93} HLA-B8 may be associated with acute inflammatory features of sarcoidosis that confer a more favorable prognosis, whereas HLA-B13 may be associated with a progressive and protracted course.⁹³

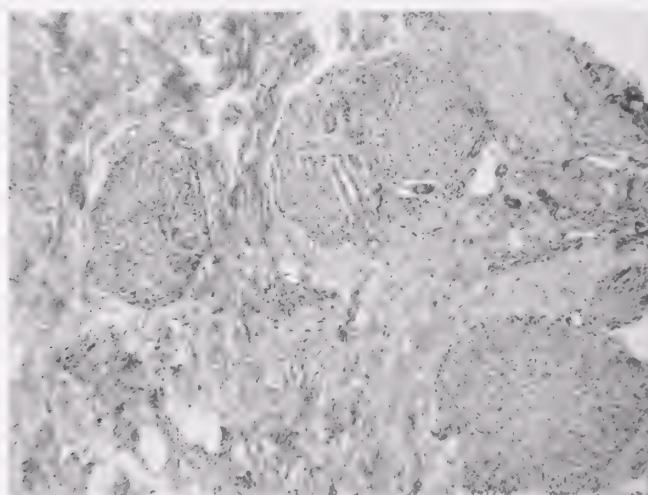


Fig 6. Kveim-Siltzbach test. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Löfgren's syndrome (BHL, erythema nodosum, polyarthritis, and fever) portends an excellent prognosis, with high rates (>85%) of spontaneous remission.^{91,94} It occurs in 30% of Caucasians with sarcoidosis but is rare in blacks.^{88,95} Clinical factors associated with a worse prognosis in sarcoidosis include age at onset > 40 years, hypercalcemia, extrathoracic disease, lupus pernio, splenomegaly, pulmonary infiltrates on chest radiographs, chronic uveitis, cystic bone lesions, nasal mucosal sarcoidosis, and lower annual family income.^{95,96} Cardiac involvement and the development of pulmonary hypertension have also been associated with a poor prognosis.

The fatality rate in sarcoidosis is estimated to be 1% to 5%.^{2,79} In the United Kingdom and the United States, most deaths are secondary to pulmonary complications.⁹⁷ In Japan, patients typically die from cardiac involvement.⁹⁸ The overall prognosis for patients with neurosarcoidosis is poorer than that for patients with disease outside the nervous system; however, certain neurologic manifestations have favorable outcomes.⁸⁴ The mortality rate from neurosarcoidosis is approximately 5% to 15%.⁹⁹

HYPERSENSITIVITY PNEUMONITIS

HP, also called extrinsic allergic alveolitis, is a complex syndrome with a variable clinical presentation and disease course. Although numerous antigenic triggers are known to incite the disease (Table 4), only a minority of subjects exposed to the antigens develop HP.

The pathogenesis of HP involves repeated antigen exposure, immunological sensitization of the host to the antigen, and an immune inflammatory response causing pulmonary and systemic symptoms. The immune response is characterized by interstitial and alveolar inflammation with granuloma formation. The latter process appears to involve a combination

of type III hypersensitivity reactions and delayed type IV hypersensitivity reactions to the inhaled antigens.

The clinical presentations of HP are classified as acute, subacute, or chronic according to the frequency, length, and intensity of exposure as well as the duration of the subsequent illness. Chronic, advanced disease is characterized by increasing fibrosis with thickening of the alveolar septa and basement membrane. Chronic HP is discussed in this review.¹⁰⁰

Diagnosis

Patients with chronic HP usually report the insidious onset of cough, dyspnea, fatigue, and weight loss. The diagnosis of chronic HP is often delayed, and inevitably, patients develop pulmonary fibrosis and progressive, irreversible respiratory insufficiency.^{101,102} In many cases, chronic HP may be indistinguishable from IPF and other forms of fibrotic lung disease.^{102,103}

The radiographic features of chronic HP may be similar to those of IPF.¹⁰³ The chest radiograph shows progressive fibrotic changes and loss of lung volume, predominately in the upper lobes. Nodular or ground-glass densities are seldom appreciated. It is important to emphasize that the chest radiograph in chronic HP may also be normal.¹⁰⁴

HRCT is more sensitive than the chest radiograph.¹⁰⁵⁻¹⁰⁷ Irregular opacities, traction bronchiectasis, loss of lung volume, and honeycombing are cardinal radiologic features of chronic HP (Figure 7).

Open or video-assisted thoracoscopic lung biopsy is usually required for diagnosis of chronic HP. The histopathologic features are not pathognomonic for HP, but characteristic findings include minimal interstitial inflammation, multinucleated giant cells in membranous and respiratory bronchioles, poorly formed, noncaseating granulomas, constrictive bronchiolitis, and moderate or severe fibrosis (Figure 8).¹⁰⁸

In certain cases, histopathology may be nondiagnostic, and chronic HP may be difficult to diagnose. Specific criteria may be used to make the diagnosis of acute or subacute HP.¹⁰⁹ Significant predictors of HP are the following: (1) exposure to a known offending antigen, (2) positive precipitating antibodies to the offending antigen, (3) recurrent episodes of symptoms, (4) inspiratory crackles on physical examination, (5) symptoms occurring 4 to 8 hours after exposure, and (6) weight loss. Exposure to the offending antigen was the strongest predictor for HP. The probability of having HP depended on the number of features present. Patients who did not

Table 4. Common Causes of Hypersensitivity Pneumonitis.

Farm worker's lung: Thermophilic actinomycetes
Winemaker's lung: <i>Botrytis cinerea</i>
Coffee worker's lung: Coffee bean dust
Lifeguard's lung: Aerosolized endotoxin
Poultry worker's lung: Avian antigens
Laboratory worker's lung: Rodent antigens
Miller's lung: Wheat weevil
Woodworker's lung: <i>Penicillium chrysogenum</i>
Detergent worker's lung: <i>Bacillus subtilis</i>
Epoxy-resin lung: Phthalic anhydride

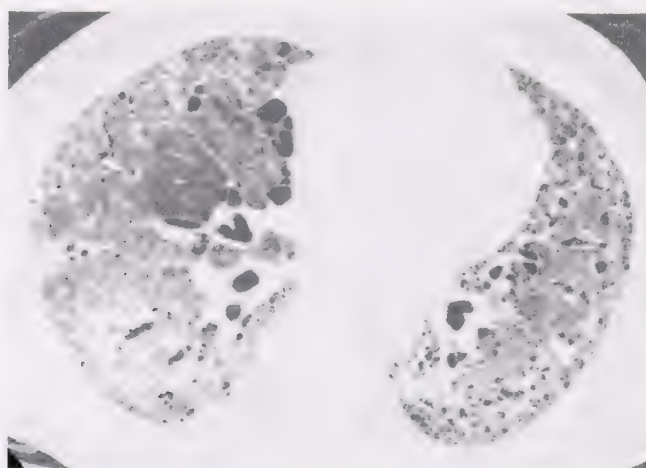


Fig 7. In the chronic stage, high-resolution computed tomography reveals irregular opacities, traction bronchiectasis, loss of lung volume, and honeycombing.



Fig 8. Interstitial inflammation and fibrosis with multiple nonnecrotizing granulomas in hypersensitivity pneumonitis. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

exhibit any of these features had a 0% chance of HP, whereas those who demonstrated all 6 criteria had a 98% probability of disease.¹¹⁰ Unfortunately, these diagnostic criteria are rarely present or attainable in chronic HP.

Treatment and Prognosis

It is important to determine the offending antigen in HP so patients can avoid the stimulus for their disease. When antigen avoidance is impractical, exposure should be limited. Most symptomatic patients should receive a trial of corticosteroids. Patients with acute farmer's lung (not chronic HP) who were given corticosteroids showed enhanced lung function in comparison with a control group.¹¹¹ The treated group did not demonstrate improved long-term survival, but it is unknown whether

corticosteroids alter the long-term course of HP. The initial dose and the duration of corticosteroid therapy in HP patients are also unknown.

Many patients remove themselves from exposure to the offending agent but, for uncertain reasons, experience progressive deterioration in lung function.¹¹² Lung transplantation is the best treatment option for patients with chronic HP refractory to medical therapy.

DIFFUSE PARENCHYMAL LUNG DISEASES OF KNOWN ASSOCIATION

Rheumatoid Arthritis–Associated Interstitial Lung Disease

RA is a common systemic inflammatory disease characterized by chronic symmetrical erosive synovitis. It

occurs in approximately 1% to 2% of the population and commonly affects patients in their fifth and sixth decades of life. The majority of afflicted patients suffer from impaired functional status, quality of life, and survival. Nearly 50% of patients with RA demonstrate extra-articular manifestations of the disease, and pulmonary involvement is common.^{113,114} ILD is the most common and most severe manifestation of lung disease.¹¹⁵

The pulmonary complications in RA are directly responsible for 10% to 20% of all mortality.^{104,105,115} Most lung disease occurs within the first 5 years after the initial diagnosis and may be a presenting manifestation in 10% to 20% of patients.¹¹⁵ The current prevalence of ILD in RA is unknown. In retrospective studies, clinically significant ILD was present in approximately 7% of subjects with RA.^{106,115} The prevalence of interstitial involvement was much greater in other studies.¹⁰⁷

What Are the Prognostic Factors Associated with the Development of Rheumatoid Arthritis Interstitial Lung Disease?

Rheumatoid arthritis interstitial lung disease (RA-ILD) is more common in men.^{106,108,116} Active or previous tobacco smoking is an independent risk factor for the development of RA, but the relationship between tobacco smoke and the development of RA-ILD is unclear.^{106,109}

Several serologic biomarkers predict the development of RA-ILD. Elevated levels of serum immunoglobulin M rheumatoid factor or anti-cyclic citrullinated peptide (anti-CCP) portend the progression of RA-ILD.^{117,118} High-titer rheumatoid factor is associated with RA-ILD.¹¹⁰ The pathogenic role of anti-CCP antibodies in RA-ILD is currently unclear. However, the antibodies are present in bronchoalveolar lavage (BAL) fluid cells, and this suggests that the lung is an antigenic source for anti-CCP antibodies.¹¹⁹

The histopathologic pattern of the ILD in RA-ILD has prognostic significance for the disease.¹²⁰ As a group, RA-ILD patients have a better prognosis than those with UIP/IPF.¹²⁰ However, RA-ILD patients who have UIP have a survival that is similar to that of matched subjects with UIP/IPF.^{121,122}

Interstitial Lung Disease in Systemic Sclerosis

SSc is a chronic connective tissue disease that is characterized by an accumulation of collagen in the skin and visceral organs. Patients suffer from a variety of symptoms that are primarily related to

the type and extent of organ involvement in the disease. Pulmonary disease is the most frequent cause of death in SSc.^{123–125} ILD and pulmonary arterial hypertension (PAH) are the most common pulmonary manifestations of SSc. We discuss only ILD associated with SSc.

ILD is the leading cause of morbidity and mortality in SSc.^{126,127} Although diffuse cutaneous scleroderma (dc SSc) is more frequently associated with ILD, ILD also occurs in patients with limited scleroderma (limited cutaneous SSc) and even in patients without any cutaneous sclerosis (SSc sine scleroderma).¹²⁸ The extent and severity of cutaneous involvement do not correlate with pulmonary involvement.¹²⁹

Histopathology

Although a variety of histopathologic patterns of ILD can occur in SSc, NSIP is the most common histopathologic pattern.^{130–132} The HRCT patterns of lung disease in patients with SSc were compared to those with idiopathic NSIP or IPF.¹³³ Patients with SSc exhibited disease that was less extensive and less coarse ($P > 0.001$) but showed a greater proportion of GGO ($P < 0.001$) than patients with IPF. The CT features closely resembled those visualized in patients with idiopathic NSIP.

Prognostic Factors

The factors that influence prognosis in SSc-ILD are uncertain. Early studies suggested an association between BAL alveolitis and lung function.¹³⁴ Recent studies that have compared BAL findings to those on HRCT have achieved variable results.^{135,136} The Scleroderma Lung Study (SLS) examined the association between GGO on HRCT with neutrophilia ($>4\%$) or eosinophilia ($>2\%$) on BAL.¹³⁷ HRCT was more sensitive than BAL in identifying active alveolitis: 90.6% versus 70%. It is important to emphasize that the interpretation of BAL differential cell counts may vary widely among institutions. Treatment decisions based solely on BAL findings should be made cautiously.

ILD on HRCT,¹³⁸ reduced baseline FVC,¹³⁹ and oxygen desaturation with exercise¹³⁷ portend a poor prognosis in patients with SSc-ILD. Serum antibodies to topoisomerase I (anti-Scl-70 antibodies) also have prognostic significance. The level of anti-Scl-70 antibodies in the sera is negatively correlated with the FVC and HRCT score.¹⁴⁰ In addition, anti-topoisomerase I-positive SSc patients who become anti-topoisomerase I-negative during the course of their disease have a more favorable

outcome (ie, less skin and lung involvement with better survival).¹⁵⁷

PAH frequently complicates SSc, and its presence is a poor prognostic factor.^{141,142} Guidelines for the treatment of PAH associated with SSc are evolving. Most clinicians extrapolate from treatment guidelines for IPH. The implications of PAH in SSc are discussed elsewhere.

Treatment

The SLS examined the safety and efficacy of oral cyclophosphamide versus placebo in patients with early SSc-ILD.¹⁴³ The Fibrosing Alveolitis in Scleroderma Trial (FAST) evaluated the effect of adding azathioprine to corticosteroids and intravenous cyclophosphamide versus placebo in SSc-ILD patients.¹⁴⁴

The SLS trial found that treatment with 1 year of oral cyclophosphamide was associated with a modest beneficial effect on lung function, dyspnea, thickening of the skin, and health-related quality of life. The mean absolute difference in the adjusted 12-month FVC percentage predicted was superior by 2.53% (95% confidence interval, 0.28–4.79) in the group of patients who received cyclophosphamide versus those who received placebo ($P < 0.03$). Those in the placebo group with more severe fibrosis on baseline HRCT had the greatest decline in FVC during the 1-year treatment period. Not surprisingly, there was a greater frequency of adverse events in the cyclophosphamide group. Cyclophosphamide did not appear to benefit patients beyond the initial year of treatment.

The FAST study demonstrated a trend toward improvement in FVC in the treatment group ($P = 0.08$). Additional studies are required to substantiate the results of the SLS and FAST.

Dermatomyositis/Polymyositis: The Antisynthetase Syndrome

The idiopathic inflammatory myopathies (IIMs) are a group of rare autoimmune disorders of unknown etiology characterized by skeletal muscle inflammation and a variety of other systemic symptoms. Classically, 3 different subtypes of IIMs have been described: DM, PM, and inclusion body myositis. DM and PM are frequently associated with pulmonary disease. The disorders are characterized and differentiated by serologic biomarkers of disease activity. The anti-Jo-1 antibody, directed against histidyl transfer RNA synthetase, is one of the best studied biomarkers.^{145,146}

Anti-Jo-1 was first identified in serum from a patient (John P) who had PM and ILD.¹⁴⁵ The

antibody was subsequently discovered in patients who exhibited a similar constellation of signs and symptoms (ILD, arthritis, Raynaud's phenomenon, and mechanic's hands), which was later called AS.¹⁴⁷ Patients with AS tend to have ILD and have a greater mortality than other IIM patients.^{134,148} To date, 7 antisynthetase antibodies have been identified. Testing for anti-Jo-1 is performed routinely. The other antibodies are tested at several specialized centers within the United States. Anti-Jo-1 has been found in 10% to 40% of patients with PM and in 2% to 10% of those with DM.^{138,140,146,149–151} The clinical manifestations of AS include PM or DM, diffuse ILD, polyarthritis, Raynaud's phenomenon, and thick cracked skin over the tips and sides of the fingers (mechanic's hands; Figure 9). ILD is the most common clinical manifestation of AS, occurring in greater than 70% of patients.^{140,152–155}

ILD in DM and PM is histopathologically indistinguishable from the type that occurs in RA and scleroderma. In autopsy studies, several histopathologic patterns have been associated with DM and PM. These include NSIP, UIP, cryptogenic organizing pneumonia, and diffuse alveolar damage (DAD).¹⁴⁸ In 1 study, NSIP most commonly complicated DM or PM.¹⁴⁸ DAD, UIP, and COP were interstitial diseases less frequently associated with DM and PM.

Prognosis in AS is primarily determined by the extent of ILD.^{134,156} In a meta-analysis of 745 patients with PM and DM, the mortality associated with ILD was approximately 40%.¹⁵⁷

The histopathologic pattern of interstitial disease also influences prognosis. In 1 study, approximately 50 patients with ILD and either PM or DM (38% of whom were Jo-1 antibody-positive) were treated with prednisone and azathioprine or methotrexate.¹⁴⁸ Survival was significantly better than that observed for historical control subjects with idiopathic UIP and



Fig 9. Mechanic's hands in a patient with antisynthetase syndrome. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

was consistent with survival previously reported for idiopathic NSIP (with a 5-year survival of approximately 60%). Mortality in Jo-1–positive patients versus Jo-1–negative patients was not statistically significant.

In another study, DM or PM patients with COP had a more favorable prognosis than those with UIP.¹⁵⁸ DM or PM patients with DAD demonstrated a poor prognosis. In addition, the histopathologic pattern of lung disease was a better predictor of survival than the radiographic features or the clinical presentation.

La Corte and colleagues¹⁵⁹ compared the clinical features and autoantibody profiles of patients with AS to those of DM/PM patients without AS. AS patients who were anti-Jo-1–positive and anti-Ro/SSA–positive had more significant ILD on HRCT. Interestingly, neither antibody positivity nor ILD on HRCT adversely affected survival.

The treatment for AS-associated ILD is similar to that which is used for other ILDs. Cyclophosphamide or tacrolimus (FK 506) is modestly effective in patients with DM/PM-associated ILD (although the antibody profile of these patients is unknown).¹³⁴ Anti-Jo-1 antibody–positive patients do not respond as well as seronegative patients to second-line treatments such as methotrexate or azathioprine.¹⁶⁰ Several case series and case reports suggest that rituximab is efficacious for the treatment of AS-related ILD.^{155,161,162} However, other novel therapies are needed to treat the ILD in AS and other CVDs.

HERMANSKY-PUDLAK SYNDROME

HPS is a rare genetic disorder of autosomal recessive inheritance and is characterized by the clinical triad of oculocutaneous albinism, bleeding diathesis, and pulmonary disease.^{163–166} HPS results from impaired intracellular trafficking of 3 membrane-bound organelles: melanosomes, platelet dense bodies, and lysosomes.¹⁶⁴ The impaired biosynthesis of melanosomes causes oculocutaneous albinism, whereas the malformation of platelet dense bodies causes bleeding diathesis. The impaired biosynthesis of lysosomes is the primary cause of the morbidity associated with HPS. Lysosomal dysfunction causes the accumulation of ceroid lipofuscin within organs, particularly the lungs, and pulmonary fibrosis.¹⁶⁴ The characteristic histopathology of HPS can be seen in Figure 10.

There are 7 known HPS genes in humans.^{165–168} The HPS gene mutation is considered to be one of the most prevalent single-gene disorders in northwest

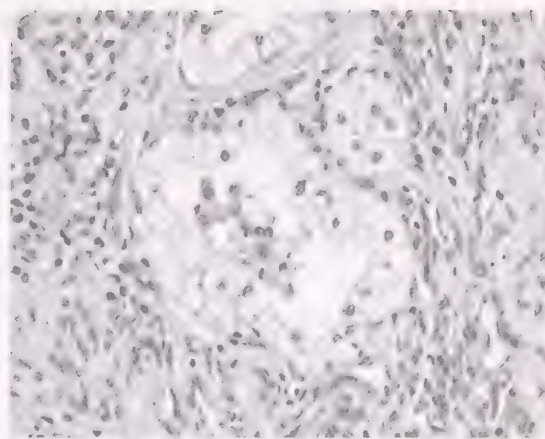


Fig 10. Biopsy specimen of Hermansky-Pudlak syndrome demonstrating collections of foamy type II pneumocytes and alveolar histiocytes laden with ceroid, an insoluble lipoprotein substance. Reprinted with permission from *Respiration*.¹⁶⁶ Copyright 2006, European Association for Bronchology and Interventional Pulmonology. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Puerto Rico, home to the largest cohort of known patients.^{108,169} The HPS1 subtype is the most common HPS gene.¹⁶⁷

Patients with the HPS1 mutation typically develop pulmonary fibrosis in their third decade of life. Pulmonary fibrosis occurs in 80% of individuals with the HPS1 genotype and causes premature death in 50% of these patients.^{170,171} Females are twice as likely to manifest pulmonary disease.¹⁷² Colitis occurs in approximately 15% of HPS1 patients.¹⁷³

The HPS2 subtype is extremely rare. HPS2 patients have mild immunodeficiency, oculocutaneous albinism, and bleeding tendencies. Pulmonary disease is uncommon and, if present, is usually mild.

Patients with HPS3 mutations exhibit mild phenotypic features of the disease. HPS4 patients have not been extensively studied. The protein products of the HPS5 and HPS6 genes probably interact in a common pathway. Patients with HPS5 or HPS6 mutations have similar phenotypes. These patients usually display oculocutaneous albinism and bleeding dyscrasia but do not exhibit pulmonary disease. To date, only 1 patient with HPS7 has been identified and studied.¹⁷⁴ The patient is a 48-year-old Portuguese female with oculocutaneous albinism, bleeding tendencies, and mild exertional dyspnea. At present, the patient's HRCT is devoid of pulmonary abnormalities.

The bleeding diathesis and pulmonary fibrosis associated with HPS are difficult to treat.¹⁶⁶ Immunosuppressive/anti-inflammatory agents have little impact on ceroid lipofuscin deposition and are

marginally effective. In addition, these drugs may have serious adverse affects.

Pirfenidone is an anti-inflammatory, antioxidant, and antifibrotic medication that may delay the onset of fibrosis. HPS1 patients who received pirfenidone showed an approximately 8%/year slower decline in pulmonary function in comparison with an untreated control group.¹⁷⁵ The results of the study require validation. The National Institutes of Health is currently enrolling patients with comorbid HS and ILD into a large clinical trial to ascertain the effects of pirfenidone on lung function.¹⁷⁶

To date, there have been 3 successful lung transplants performed in patients with HPS, and there are additional patients who are being considered for the procedure.¹⁶⁶ Lung transplantation is not a first-line treatment for patients with HPS. However, it is a viable treatment option for patients with end-stage lung disease for whom medical therapy has failed.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Katzenstein AA, Askin FB. Surgical pathology of non-neoplastic lung disease. *Major Probl Pathol* 1982; 13: 1–430.
- Demedts M, Costabel U. ATS/ERS international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *Eur Respir J* 2002; 19: 794–796.
- Trulock EP. Lung transplantation. *Am J Respir Crit Care Med* 1997; 155: 789–818.
- Katzenstein AL, Myers JL. Idiopathic pulmonary fibrosis: clinical relevance of pathologic classification. *Am J Respir Crit Care Med* 1998; 157(pt 1): 1301–1315.
- Hunninghake GW, Lynch DA, Galvin JR, et al. Radiologic findings are strongly associated with a pathologic diagnosis of usual interstitial pneumonia. *Chest* 2003; 124: 1215–1223.
- Daniil ZD, Gilchrist FC, Nicholson AG, et al. A histologic pattern of nonspecific interstitial pneumonia is associated with a better prognosis than usual interstitial pneumonia in patients with cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 1999; 160: 899–905.
- Cottin V, Donsbeck AV, Revel D, et al. Nonspecific interstitial pneumonia. Individualization of a clinicopathologic entity in a series of 12 patients. *Am J Respir Crit Care Med* 1998; 158: 1286–1293.
- Nicholson AG, Colby TV, du Bois RM, et al. The prognostic significance of the histologic pattern of interstitial pneumonia in patients presenting with the clinical entity of cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 2000; 162: 2213–2217.
- Travis WD, Matsui K, Moss J, Ferrans VJ. Idiopathic nonspecific interstitial pneumonia: prognostic significance of cellular and fibrosing patterns: survival comparison with usual interstitial pneumonia and desquamative interstitial pneumonia. *Am J Surg Pathol* 2000; 24: 19–33.
- Fujita J, Yamadori I, Suemitsu I, et al. Clinical features of non-specific interstitial pneumonia. *Respir Med* 1999; 93: 113–118.
- Flaherty KR, Martinez FJ, Travis W, Lynch JP III. Nonspecific interstitial pneumonia (NSIP). *Semin Respir Crit Care Med* 2001; 22: 423–434.
- Flaherty KR, Travis WD, Colby TV, et al. Histopathologic variability in usual and nonspecific interstitial pneumonias. *Am J Respir Crit Care Med* 2001; 164: 1722–1727.
- Kinder BW, Collard HR, Koth L, et al. Idiopathic nonspecific interstitial pneumonia: lung manifestation of undifferentiated connective tissue disease? *Am J Respir Crit Care Med* 2007; 176: 691–697.
- Flaherty KR, Toews GB, Travis WD, et al. Clinical significance of histological classification of idiopathic interstitial pneumonia. *Eur Respir J* 2002; 19: 275–283.
- Flaherty KR, Thwaite EL, Kazerooni EA, et al. Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax* 2003; 58: 143–148.
- Katzenstein AL, Fiorelli RF. Nonspecific interstitial pneumonia/fibrosis. Histologic features and clinical significance. *Am J Surg Pathol* 1994; 18: 136–147.
- Hunninghake GW, Zimmerman MB, Schwartz DA, et al. Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2001; 164: 193–196.
- Hartman TE, Swensen SJ, Hansell DM, et al. Nonspecific interstitial pneumonia: variable appearance at high-resolution chest CT. *Radiology* 2000; 217: 701–705.
- MacDonald SL, Rubens MB, Hansell DM, et al. Nonspecific interstitial pneumonia and usual interstitial pneumonia: comparative appearances at and diagnostic accuracy of thin-section CT. *Radiology* 2001; 221: 600–605.
- Qureshi RA, Ahmed TA, Grayson AD, et al. Does lung biopsy help patients with interstitial lung disease? *Eur J Cardiothorac Surg* 2002; 21: 621–626; discussion 626.
- Arase Y, Ikeda K, Tsubota A, et al. Usefulness of serum KL-6 for early diagnosis of idiopathic pulmonary fibrosis in patients with hepatitis C virus. *Hepatol Res* 2003; 27: 89–94.
- Ashitani J, Mukae H, Taniguchi H, et al. Granulocyte-colony stimulating factor levels in bronchoalveolar lavage fluid from patients with idiopathic pulmonary fibrosis. *Thorax* 1999; 54: 1015–1020.
- Capelli A, Lusuuardi M, Cerutti CG, Donner CF. Lung alkaline phosphatase as a marker of fibrosis in chronic interstitial disorders. *Am J Respir Crit Care Med* 1997; 155: 249–253.
- Dobashi N, Fujita J, Ohtsuki Y, et al. Elevated serum and BAL cytokeratin 19 fragment in pulmonary fibrosis and acute interstitial pneumonia. *Eur Respir J* 1999; 14: 574–578.
- Fujita J, Obayashi Y, Yamadori I, et al. Marked elevation of CA19–9 in a patient with idiopathic

- pulmonary fibrosis: CA19-9 as a bad prognostic factor. *Respirology* 1998; 3: 211–214.
26. Yamanouchi H, Fujita J, Yoshinouchi T, et al. Measurement of hepatocyte growth factor in serum and bronchoalveolar lavage fluid in patients with pulmonary fibrosis. *Respir Med* 1998; 92: 273–278.
 27. Greene KE, King TE Jr, Kuroki Y, et al. Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. *Eur Respir J* 2002; 19: 439–446.
 28. Kadota J, Mizunoe S, Mito K, et al. High plasma concentrations of osteopontin in patients with interstitial pneumonia. *Respir Med* 2005; 99: 111–117.
 29. Kasuga I, Yonemaru M, Kiyokawa H, et al. Clinical evaluation of serum type IV collagen 7S in idiopathic pulmonary fibrosis. *Respirology* 1996; 1: 277–281.
 30. Honda Y, Kuroki Y, Shijubo N, et al. Aberrant appearance of lung surfactant protein A in sera of patients with idiopathic pulmonary fibrosis and its clinical significance. *Respiration* 1995; 62: 64–69.
 31. Kohno N, Awaya Y, Oyama T, et al. KL-6, a mucin-like glycoprotein, in bronchoalveolar lavage fluid from patients with interstitial lung disease. *Am Rev Respir Dis* 1993; 148: 637–642.
 32. Honda Y, Kuroki Y, Matsuura E, et al. Pulmonary surfactant protein D in sera and bronchoalveolar lavage fluids. *Am J Respir Crit Care Med* 1995; 152(pt 1): 1860–1866.
 33. Kubo H. Anticoagulant therapy for idiopathic pulmonary fibrosis. *Rev Port Pneumol* 2005; 11: 591–592.
 34. Matsui T, Inui N, Suda T, Chida K. Anti-endothelial cell antibodies in patients with interstitial lung diseases. *Respir Med* 2008; 102: 128–133.
 35. Nadrous HF, Pellikka PA, Krowka MJ, et al. Pulmonary hypertension in patients with idiopathic pulmonary fibrosis. *Chest* 2005; 128: 2393–2399.
 36. Nadrous HF, Pellikka PA, Krowka MJ, et al. The impact of pulmonary hypertension on survival in patients with idiopathic pulmonary fibrosis. *Chest* 2005; 128(suppl): 616S–617S.
 37. Lettieri CJ, Nathan SD, Barnett SD, et al. Prevalence and outcomes of pulmonary arterial hypertension in advanced idiopathic pulmonary fibrosis. *Chest* 2006; 129: 746–752.
 38. King TE Jr, Tooze JA, Schwarz MI, et al. Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med* 2001; 164: 1171–1181.
 39. Parambil JG, Myers JL, Ryu JH. Histopathologic features and outcome of patients with acute exacerbation of idiopathic pulmonary fibrosis undergoing surgical lung biopsy. *Chest* 2005; 128: 3310–3315.
 40. Kondoh Y, Taniguchi H, Kawabata Y, et al. Acute exacerbation in idiopathic pulmonary fibrosis. Analysis of clinical and pathologic findings in three cases. *Chest* 1993; 103: 1808–1812.
 41. Akira M, Hamada H, Sakatani M, et al. CT findings during phase of accelerated deterioration in patients with idiopathic pulmonary fibrosis. *AJR Am J Roentgenol* 1997; 168: 79–83.
 42. Ambrosini V, Cancellieri A, Chilosi M, et al. Acute exacerbation of idiopathic pulmonary fibrosis: report of a series. *Eur Respir J* 2003; 22: 821–826.
 43. Kim DS, Park JH, Park BK, et al. Acute exacerbation of idiopathic pulmonary fibrosis: frequency and clinical features. *Eur Respir J* 2006; 27: 143–150.
 44. Okamoto T, Ichiyasu H, Ichikado K, et al. Clinical analysis of the acute exacerbation in patients with idiopathic pulmonary fibrosis. *Nihon Kokyuki Gakkai Zasshi* 2006; 44: 359–367.
 45. Al-Hameed FM, Sharma S. Outcome of patients admitted to the intensive care unit for acute exacerbation of idiopathic pulmonary fibrosis. *Can Respir J* 2004; 11: 117–122.
 46. Blivet S, Philit F, Sab JM, et al. Outcome of patients with idiopathic pulmonary fibrosis admitted to the ICU for respiratory failure. *Chest* 2001; 120: 209–212.
 47. Kondoh Y, Taniguchi H, Kitaichi M, et al. Acute exacerbation of interstitial pneumonia following surgical lung biopsy. *Respir Med* 2006; 100: 1753–1759.
 48. Azuma A, Nukiwa T, Tsuboi E, et al. Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2005; 171: 1040–1047.
 49. Nishiyama O, Taniguchi H, Kondoh Y, et al. Dyspnoea at 6-min walk test in idiopathic pulmonary fibrosis: comparison with COPD. *Respir Med* 2007; 101: 833–838.
 50. Noble PW. Idiopathic pulmonary fibrosis: natural history and prognosis. *Clin Chest Med* 2006; 27(suppl 1): S11–S16.
 51. Jeon K, Chung MP, Lee KS, et al. Prognostic factors and causes of death in Korean patients with idiopathic pulmonary fibrosis. *Respir Med* 2006; 100: 451–457.
 52. Martinez FJ, Safran S, Weycker D, et al. The clinical course of patients with idiopathic pulmonary fibrosis. *Ann Intern Med* 2005; 142(pt 1): 963–967.
 53. Schwartz DA, Helmers RA, Galvin JR, et al. Determinants of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1994; 149(pt 1): 450–454.
 54. Collard HR, King TE Jr, Bartelson BB, et al. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003; 168: 538–542.
 55. King TE Jr, Schwarz MI, Brown K, et al. Idiopathic pulmonary fibrosis: relationship between histopathologic features and mortality. *Am J Respir Crit Care Med* 2001; 164: 1025–1032.
 56. Lederer DJ, Arcasoy SM, Wilt JS, et al. Six-minute-walk distance predicts waiting list survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2006; 174: 659–664.
 57. Nicholson AG, Fulford LG, Colby TV, et al. The relationship between individual histologic features and disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2002; 166: 173–177.
 58. Enomoto N, Suda T, Kato M, et al. Quantitative analysis of fibroblastic foci in usual interstitial pneumonia. *Chest* 2006; 130: 22–29.
 59. Cool CD, Groshong SD, Rai PR, et al. Fibroblast foci are not discrete sites of lung injury or repair: the fibroblast reticulum. *Am J Respir Crit Care Med* 2006; 174: 654–658.
 60. Billiau A, Heremans H, Vermeire K, Matthys P. Immunomodulatory properties of interferon-gamma. An update. *Ann N Y Acad Sci* 1998; 856: 22–32.

61. Antoniou KM, Pataka A, Bouros D, Siafakas NM. Pathogenetic pathways and novel pharmacotherapeutic targets in idiopathic pulmonary fibrosis. *Pulm Pharmacol Ther* 2007; 20: 453–461.
62. Liles WC. Immunomodulatory approaches to augment phagocyte-mediated host defense for treatment of infectious diseases. *Semin Respir Infect* 2001; 16: 11–17.
63. Giaid A, Michel RP, Stewart DJ, et al. Expression of endothelin-1 in lungs of patients with cryptogenic fibrosing alveolitis. *Lancet* 1993; 341: 1550–1554.
64. Ziesche R, Hofbauer E, Wittmann K, et al. A preliminary study of long-term treatment with interferon gamma-1b and low-dose prednisolone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 1999; 341: 1264–1269.
65. Rubin LJ, Badesch DB, Barst RJ, et al. Bosentan therapy for pulmonary arterial hypertension. *N Engl J Med* 2002; 346: 896–903.
66. Cantin AM, Hubbard RC, Crystal RG. Glutathione deficiency in the epithelial lining fluid of the lower respiratory tract in idiopathic pulmonary fibrosis. *Am Rev Respir Dis* 1989; 139: 370–372.
67. Kapanci Y, Desmouliere A, Pache JC, et al. Cytoskeletal protein modulation in pulmonary alveolar myofibroblasts during idiopathic pulmonary fibrosis. Possible role of transforming growth factor beta and tumor necrosis factor alpha. *Am J Respir Crit Care Med* 1995; 152(pt 1): 2163–2169.
68. Piguet PF, Vesin C. Treatment by human recombinant soluble TNF receptor of pulmonary fibrosis induced by bleomycin or silica in mice. *Eur Respir J* 1994; 7: 515–518.
69. Zhang Y, Lee TC, Guillemin B, et al. Enhanced IL-1 beta and tumor necrosis factor-alpha release and messenger RNA expression in macrophages from idiopathic pulmonary fibrosis or after asbestos exposure. *J Immunol* 1993; 150: 4188–4196.
70. Whyte M, Hubbard R, Meliconi R, et al. Increased risk of fibrosing alveolitis associated with interleukin-1 receptor antagonist and tumor necrosis factor-alpha gene polymorphisms. *Am J Respir Crit Care Med* 2000; 162(pt 1): 755–758.
71. Nathan SD. Therapeutic management of idiopathic pulmonary fibrosis: an evidence-based approach. *Clin Chest Med* 2006; 27(suppl 1): S27–S35.
72. Trulock EP, Edwards LB, Taylor DO, et al. The Registry of the International Society for Heart and Lung Transplantation: twentieth official adult lung and heart-lung transplant report—2003. *J Heart Lung Transplant* 2003; 22: 625–635.
73. Mason DP, Brizzio ME, Alster JM, et al. Lung transplantation for idiopathic pulmonary fibrosis. *Ann Thorac Surg* 2007; 84: 1121–1128.
74. Meyer DM, Edwards LB, Torres F, et al. Impact of recipient age and procedure type on survival after lung transplantation for pulmonary fibrosis. *Ann Thorac Surg* 2005; 79: 950–957; discussion 957–958.
75. Grgic A, Lausberg H, Heinrich M, et al. Progression of fibrosis in usual interstitial pneumonia: serial evaluation of the native lung after single lung transplantation. *Respiration* 2008; 76: 139–145.
76. Charman SC, Sharples LD, McNeil KD, Wallwork J. Assessment of survival benefit after lung transplantation by patient diagnosis. *J Heart Lung Transplant* 2002; 21: 226–232.
77. De Meester J, Smits JM, Persijn GG, Haverich A. Listing for lung transplantation: life expectancy and transplant effect, stratified by type of end-stage lung disease, the Eurotransplant experience. *J Heart Lung Transplant* 2001; 20: 518–524.
78. Newman LS, Rose CS, Maier LA. Sarcoidosis. *N Engl J Med* 1997; 336: 1224–1234.
79. Statement on sarcoidosis. Joint statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* 1999; 160: 736–755.
80. Johns CJ, Michele TM. The clinical management of sarcoidosis. A 50-year experience at the Johns Hopkins Hospital. *Medicine (Baltimore)* 1999; 78: 65–111.
81. Baughman RP, Teirstein AS, Judson MA, et al. Clinical characteristics of patients in a case control study of sarcoidosis. *Am J Respir Crit Care Med* 2001; 164(pt 1): 1885–1889.
82. Lynch JP III, Kazerooni EA, Gay SE. Pulmonary sarcoidosis. *Clin Chest Med* 1997; 18: 755–785.
83. Romer FK. Presentation of sarcoidosis and outcome of pulmonary changes. *Dan Med Bull* 1982; 29: 27–32.
84. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357: 2153–2165.
85. Scadding JG. Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. *Br Med J* 1961; 2: 1165–1172.
86. Hillerdal G, Nöu E, Osterman K, Schmekel B. Sarcoidosis: epidemiology and prognosis. A 15-year European study. *Am Rev Respir Dis* 1984; 130: 29–32.
87. Johns CJ, Schonfeld SA, Scott PP, et al. Longitudinal study of chronic sarcoidosis with low-dose maintenance corticosteroid therapy. Outcome and complications. *Ann N Y Acad Sci* 1986; 465: 702–712.
88. Gottlieb JE, Israel HL, Steiner RM, et al. Outcome in sarcoidosis. The relationship of relapse to corticosteroid therapy. *Chest* 1997; 111: 623–631.
89. Siltzbach LE, James DG, Neville E, et al. Course and prognosis of sarcoidosis around the world. *Am J Med* 1974; 57: 847–852.
90. Siltzbach LE. Sarcoidosis: clinical features and management. *Med Clin North Am* 1967; 51: 483–502.
91. Judson MA, Baughman RP, Thompson BW, et al. Two year prognosis of sarcoidosis: the ACCESS experience. *Sarcoidosis Vasc Diffuse Lung Dis* 2003; 20: 204–211.
92. Grunewald J, Eklund A, Olerup O. Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. *Am J Respir Crit Care Med* 2004; 169: 696–702.
93. Lenhart K, Kolek V, Bartova A. HLA antigens associated with sarcoidosis. *Dis Markers* 1990; 8: 23–29.
94. Gran JT, Bohmer E. Acute sarcoid arthritis: a favourable outcome? A retrospective survey of 49 patients with review of the literature. *Scand J Rheumatol* 1996; 25: 70–73.
95. Israel HL, Karlin P, Menduke H, DeLisser OG. Factors affecting outcome of sarcoidosis. Influence of race, extrathoracic involvement, and initial radiologic lung lesions. *Ann N Y Acad Sci* 1986; 465: 609–618.

96. Neville E, Walker AN, James DG. Prognostic factors predicting the outcome of sarcoidosis: an analysis of 818 patients. *Q J Med* 1983; 52: 525–533.
97. Gideon NM, Mannino DM. Sarcoidosis mortality in the United States 1979–1991: an analysis of multiple-cause mortality data. *Am J Med* 1996; 100: 423–427.
98. Iwai K, Takemura T, Kitaichi M, et al. Pathological studies on sarcoidosis autopsy. II. Early change, mode of progression and death pattern. *Acta Pathol Jpn* 1993; 43: 377–385.
99. Lury KM, Smith JK, Matheus MG, Castillo M. Neurosarcoidosis—review of imaging findings. *Semin Roentgenol* 2004; 39: 495–504.
100. Fink JN, Sosman AJ, Barboriak JJ, et al. Pigeon breeders' disease. A clinical study of a hypersensitivity pneumonitis. *Ann Intern Med* 1968; 68: 1205–1219.
101. Vourlekis JS, Schwarz MI, Cherniack RM, et al. The effect of pulmonary fibrosis on survival in patients with hypersensitivity pneumonitis. *Am J Med* 2004; 116: 662–668.
102. Buschman DL, Gamsu G, Waldron JA Jr, et al. Chronic hypersensitivity pneumonitis: use of CT in diagnosis. *AJR Am J Roentgenol* 1992; 159: 957–960.
103. Lynch DA, Newell JD, Logan PM, et al. Can CT distinguish hypersensitivity pneumonitis from idiopathic pulmonary fibrosis? *AJR Am J Roentgenol* 1995; 165: 807–811.
104. Minaur NJ, Jacoby RK, Cosh JA, et al. Outcome after 40 years with rheumatoid arthritis: a prospective study of function, disease activity, and mortality. *J Rheumatol Suppl* 2004; 69: 3–8.
105. Sihvonen S, Korpela M, Laippala P, et al. Death rates and causes of death in patients with rheumatoid arthritis: a population-based study. *Scand J Rheumatol* 2004; 33: 221–227.
106. Tureson C, O'Fallon WM, Crowson CS, et al. Extra-articular disease manifestations in rheumatoid arthritis: incidence trends and risk factors over 46 years. *Ann Rheum Dis* 2003; 62: 722–727.
107. Zrour SH, Touzi M, Bejia I, et al. Correlations between high-resolution computed tomography of the chest and clinical function in patients with rheumatoid arthritis. Prospective study in 75 patients. *Joint Bone Spine* 2005; 72: 41–47.
108. Pérez-Padilla R, Gaxiola M, Salas J, et al. Bronchiolitis in chronic pigeon breeder's disease. Morphologic evidence of a spectrum of small airway lesions in hypersensitivity pneumonitis induced by avian antigens. *Chest* 1996; 110: 371–377.
109. Saag KG, Kolluri S, Koehnke RK, et al. Rheumatoid arthritis lung disease. Determinants of radiographic and physiologic abnormalities. *Arthritis Rheum* 1996; 39: 1711–1719.
110. Lacasse Y, Selman M, Costabel U, et al. Clinical diagnosis of hypersensitivity pneumonitis. *Am J Respir Crit Care Med* 2003; 168: 952–958.
111. Kokkarinen JJ, Tukiainen HO, Terho EO. Effect of corticosteroid treatment on the recovery of pulmonary function in farmer's lung. *Am Rev Respir Dis* 1992; 145: 3–5.
112. Barbee RA, Callies Q, Dickie HA, Rankin J. The long-term prognosis in farmer's lung. *Am Rev Respir Dis* 1968; 97: 223–231.
113. Carmona L, González-Alvaro I, Balsa A, et al. Rheumatoid arthritis in Spain: occurrence of extra-articular manifestations and estimates of disease severity. *Ann Rheum Dis* 2003; 62: 897–900.
114. Kim DS. Interstitial lung disease in rheumatoid arthritis: recent advances. *Curr Opin Pulm Med* 2006; 12: 346–353.
115. Brown KK. Rheumatoid lung disease. *Proc Am Thorac Soc* 2007; 4: 443–448.
116. Gabbay E, Tarala R, Will R, et al. Interstitial lung disease in recent onset rheumatoid arthritis. *Am J Respir Crit Care Med* 1997; 156(pt 1): 528–535.
117. Manfredsdottir VF, Vikingsdottir T, Jonsson T, et al. The effects of tobacco smoking and rheumatoid factor seropositivity on disease activity and joint damage in early rheumatoid arthritis. *Rheumatology (Oxford)* 2006; 45: 734–740.
118. Masdottir B, Jónsson T, Manfredsdottir V, et al. Smoking, rheumatoid factor isotypes and severity of rheumatoid arthritis. *Rheumatology (Oxford)* 2000; 39: 1202–1205.
119. Klareskog L, Stolt P, Lundberg K, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 2006; 54: 38–46.
120. Park JH, Kim DS, Park IN, et al. Prognosis of fibrotic interstitial pneumonia: idiopathic versus collagen vascular disease-related subtypes. *Am J Respir Crit Care Med* 2007; 175: 705–711.
121. Lee HK, Kim DS, Yoo B, et al. Histopathologic pattern and clinical features of rheumatoid arthritis-associated interstitial lung disease. *Chest* 2005; 127: 2019–2027.
122. Yousem SA, Colby TV, Carrington CB. Lung biopsy in rheumatoid arthritis. *Am Rev Respir Dis* 1985; 131: 770–777.
123. Simeon CP, Armadans L, Fonollosa V, et al. Survival prognostic factors and markers of morbidity in Spanish patients with systemic sclerosis. *Ann Rheum Dis* 1997; 56: 723–728.
124. Cossio M, Menon Y, Wilson W, de Boishblanc BP. Life-threatening complications of systemic sclerosis. *Crit Care Clin* 2002; 18: 819–839.
125. Ostojic P, Cerinic MM, Silver R, et al. Interstitial lung disease in systemic sclerosis. *Lung* 2007; 185: 211–220.
126. Altman RD, Medsger TA Jr, Bloch DA, Michel BA. Predictors of survival in systemic sclerosis (scleroderma). *Arthritis Rheum* 1991; 34: 403–413.
127. Steen VD, Medsger TA Jr. Severe organ involvement in systemic sclerosis with diffuse scleroderma. *Arthritis Rheum* 2000; 43: 2437–2444.
128. Ostojic P, Damjanov N. Different clinical features in patients with limited and diffuse cutaneous systemic sclerosis. *Clin Rheumatol* 2006; 25: 453–457.
129. Morelli S, Barbieri C, Sgreccia A, et al. Relationship between cutaneous and pulmonary involvement in systemic sclerosis. *J Rheumatol* 1997; 24: 81–85.
130. Kim DS, Yoo B, Lee JS, et al. The major histopathologic pattern of pulmonary fibrosis in scleroderma is nonspecific interstitial pneumonia. *Sarcoidosis Vasc Diffuse Lung Dis* 2002; 19: 121–127.
131. Bouros D, Wells AU, Nicholson AG, et al. Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship

- to outcome. *Am J Respir Crit Care Med* 2002; 165: 1581–1586.
132. Fujita J, Yoshinouchi T, Ohtsuki Y, et al. Non-specific interstitial pneumonia as pulmonary involvement of systemic sclerosis. *Ann Rheum Dis* 2001; 60: 281–283.
 133. Desai SR, Veeraraghavan S, Hansell DM, et al. CT features of lung disease in patients with systemic sclerosis: comparison with idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia. *Radiology* 2004; 232: 560–567.
 134. Imbert-Masseau A, Hamidou M, Agard C, et al. Antisynthetase syndrome. *Joint Bone Spine* 2003; 70: 161–168.
 135. Remy-Jardin M, Remy J, Wallaert B, et al. Pulmonary involvement in progressive systemic sclerosis: sequential evaluation with CT, pulmonary function tests, and bronchoalveolar lavage. *Radiology* 1993; 188: 499–506.
 136. De Santis M, Bosello S, La Torre G, et al. Functional, radiological and biological markers of alveolitis and infections of the lower respiratory tract in patients with systemic sclerosis. *Respir Res* 2005; 6: 96.
 137. Highland KB, Garin MC, Brown KK. The spectrum of scleroderma lung disease. *Semin Respir Crit Care Med* 2007; 28: 418–429.
 138. Vazquez-Abad D, Rothfield NF. Sensitivity and specificity of anti-Jo-1 antibodies in autoimmune diseases with myositis. *Arthritis Rheum* 1996; 39: 292–296.
 139. Plastiras SC, Karadimitrakakis SP, Ziakas PD, et al. Scleroderma lung: initial forced vital capacity as predictor of pulmonary function decline. *Arthritis Rheum* 2006; 55: 598–602.
 140. Love LA, Leff RL, Fraser DD, et al. A new approach to the classification of idiopathic inflammatory myopathy: myositis-specific autoantibodies define useful homogeneous patient groups. *Medicine (Baltimore)* 1991; 70: 360–374.
 141. McLaughlin VV, Presberg KW, Doyle RL, et al. Prognosis of pulmonary arterial hypertension: ACCP evidence-based clinical practice guidelines. *Chest* 2004; 126(suppl): 78S–92S.
 142. Kawut SM, Taichman DB, Archer-Chicko CL, et al. Hemodynamics and survival in patients with pulmonary arterial hypertension related to systemic sclerosis. *Chest* 2003; 123: 344–350.
 143. Tashkin DP, Elashoff R, Clements PJ, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006; 354: 2655–2666.
 144. Hoyles RK, Ellis RW, Wellsbury J, et al. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis Rheum* 2006; 54: 3962–3970.
 145. Wasicek CA, Reichlin M, Montes M, Raghu G. Polymyositis and interstitial lung disease in a patient with anti-Jo1 prototype. *Am J Med* 1984; 76: 538–544.
 146. Nishikai M, Reichlin M. Heterogeneity of precipitating antibodies in polymyositis and dermatomyositis. Characterization of the Jo-1 antibody system. *Arthritis Rheum* 1980; 23: 881–888.
 147. Marguerie C, Bunn CC, Beynon HL, et al. Polymyositis, pulmonary fibrosis and autoantibodies to aminoacyl-tRNA synthetase enzymes. *Q J Med* 1990; 77: 1019–1038.
 148. Douglas WW, Tazelaar HD, Hartman TE, et al. Polymyositis-dermatomyositis-associated interstitial lung disease. *Am J Respir Crit Care Med* 2001; 164: 1182–1185.
 149. Yoshida S, Akizuki M, Mimori T, et al. The precipitating antibody to an acidic nuclear protein antigen, the Jo-1, in connective tissue diseases. A marker for a subset of polymyositis with interstitial pulmonary fibrosis. *Arthritis Rheum* 1983; 26: 604–611.
 150. Reichlin M, Arnett FC Jr. Multiplicity of antibodies in myositis sera. *Arthritis Rheum* 1984; 27: 1150–1156.
 151. Dickey BF, Myers AR. Pulmonary disease in polymyositis/dermatomyositis. *Semin Arthritis Rheum* 1984; 14: 60–76.
 152. Salmeron G, Greenberg SD, Lidsky MD. Polymyositis and diffuse interstitial lung disease. A review of the pulmonary histopathologic findings. *Arch Intern Med* 1981; 141: 1005–1010.
 153. Tymms KE, Webb J. Dermatopolymyositis and other connective tissue diseases: a review of 105 cases. *J Rheumatol* 1985; 12: 1140–1148.
 154. Marie I, Hachulla E, Chérin P, et al. Interstitial lung disease in polymyositis and dermatomyositis. *Arthritis Rheum* 2002; 47: 614–622.
 155. Lambotte O, Kotb R, Maigne G, et al. Efficacy of rituximab in refractory polymyositis. *J Rheumatol* 2005; 32: 1369–1370.
 156. Tubéry M, Lauque D, Murris M, et al. Course of interstitial lung diseases associated with inflammatory myopathies. Apropos of 5 cases. *Ann Med Interne (Paris)* 1997; 148: 2–10.
 157. Arsura EL, Greenberg AS. Adverse impact of interstitial pulmonary fibrosis on prognosis in polymyositis and dermatomyositis. *Semin Arthritis Rheum* 1988; 18: 29–37.
 158. Tazelaar HD, Viggiano RW, Pickersgill J, Colby TV. Interstitial lung disease in polymyositis and dermatomyositis. Clinical features and prognosis as correlated with histologic findings. *Am Rev Respir Dis* 1990; 141: 727–733.
 159. La Corte R, Lo Mo Naco A, Locaputo A, et al. In patients with antisynthetase syndrome the occurrence of anti-Ro/SSA antibodies causes a more severe interstitial lung disease. *Autoimmunity* 2006; 39: 249–253.
 160. Joffe MM, Love LA, Leff RL, et al. Drug therapy of the idiopathic inflammatory myopathies: predictors of response to prednisone, azathioprine, and methotrexate and a comparison of their efficacy. *Am J Med* 1993; 94: 379–387.
 161. Levine TD. Rituximab in the treatment of dermatomyositis: an open-label pilot study. *Arthritis Rheum* 2005; 52: 601–607.
 162. Noss EH, Hausner-Sypek DL, Weinblatt ME. Rituximab as therapy for refractory polymyositis and dermatomyositis. *J Rheumatol* 2006; 33: 1021–1026.
 163. Bachli EB, Brack T, Eppler E, et al. Hermansky-Pudlak syndrome type 4 in a patient from Sri Lanka with pulmonary fibrosis. *Am J Med Genet A* 2004; 127: 201–207.
 164. Boissy RE, Richmond B, Huizing M, et al. Melanocyte-specific proteins are aberrantly trafficked in

- melanocytes of Hermansky-Pudlak syndrome-type 3. *Am J Pathol* 2005; 166: 231–240.
165. Poletti V, Costabel U, Casoni GL, et al. Rare infiltrative lung diseases: a challenge for clinicians. *Respiration* 2004; 71: 431–443.
166. Pierson DM, Ionescu D, Qing G, et al. Pulmonary fibrosis in Hermansky-Pudlak syndrome. A case report and review. *Respiration* 2006; 73: 382–395.
167. Oh J, Bailin T, Fukai K, et al. Positional cloning of a gene for Hermansky-Pudlak syndrome, a disorder of cytoplasmic organelles. *Nat Genet* 1996; 14: 300–306.
168. Anikster Y, Huizing M, White J, et al. Mutation of a new gene causes a unique form of Hermansky-Pudlak syndrome in a genetic isolate of central Puerto Rico. *Nat Genet* 2001; 28: 376–380.
169. Wildenberg SC, Oetting WS, Almodóvar C, et al. A gene causing Hermansky-Pudlak syndrome in a Puerto Rican population maps to chromosome 10q2. *Am J Hum Genet* 1995; 57: 755–765.
170. Shotelersuk V, Dell'Angelica EC, Hartnell L, et al. A new variant of Hermansky-Pudlak syndrome due to mutations in a gene responsible for vesicle formation. *Am J Med* 2000; 108: 423–427.
171. Anderson PD, Huizing M, Claassen DA, et al. Hermansky-Pudlak syndrome type 4 (HPS-4): clinical and molecular characteristics. *Hum Genet* 2003; 113: 10–17.
172. Parker MS, Rosado Shipley W, de Christenson ML, et al. The Hermansky-Pudlak syndrome. *Ann Diagn Pathol* 1997; 1: 99–103.
173. Gahl WA, Brantly M, Kaiser-Kupfer MI, et al. Genetic defects and clinical characteristics of patients with a form of oculocutaneous albinism (Hermansky-Pudlak syndrome). *N Engl J Med* 1998; 338: 1258–1264.
174. Li W, Zhang Q, Oiso N, et al. Hermansky-Pudlak syndrome type 7 (HPS-7) results from mutant dysbindin, a member of the biogenesis of lysosome-related organelles complex 1 (BLOC-1). *Nat Genet* 2003; 35: 84–89.
175. Gahl WA, Brantly M, Troendle J, et al. Effect of pirfenidone on the pulmonary fibrosis of Hermansky-Pudlak syndrome. *Mol Genet Metab* 2002; 76: 234–242.
176. McKane BW, Fernandez F, Narayanan K, et al. Pirfenidone inhibits obliterative airway disease in a murine heterotopic tracheal transplant model. *Transplantation* 2004; 77: 664–669.
177. Visscher DW, Myers JL. Histologic spectrum of idiopathic interstitial pneumonias. *Proc Am Thorac Soc* 2006; 3: 322–329.
178. du Bois R, King TE Jr. Challenges in pulmonary fibrosis x 5: the NSIP/UIP debate. *Thorax* 2007; 62: 1008–1012.

Gastroesophageal Reflux Disease and Idiopathic Pulmonary Fibrosis

Yevgenia Y. Pashinsky, MD,¹ Barry W. Jaffin, MD,¹ and Virginia R. Litle, MD²

¹Division of Gastroenterology, Mount Sinai Medical Center, New York, NY

²Division of Thoracic and Foregut Surgery, University of Rochester Medical Center, Rochester, NY

ABSTRACT

Gastroesophageal reflux disease occurs with a higher prevalence in patients with idiopathic pulmonary fibrosis than in matched controls. Silent reflux occurs in about a third of patients with significant gastroesophageal reflux disease; thus, objective measurements are required to evaluate gastroesophageal reflux disease in patients with advanced lung diseases. We provide here a detailed description of acid and non-acid reflux and the diagnostic evaluation for pulmonologists and lung transplant surgeons suspecting reflux as a contributing factor in advanced lung diseases. We review the evidence for gastroesophageal reflux disease causing idiopathic pulmonary fibrosis and other select pulmonary diseases and the potential role of antireflux surgery in the management of advanced lung disease and transplant patients. *Mt Sinai J Med* 76:24–29, 2009. © 2009 Mount Sinai School of Medicine

Key Words: advanced lung disease, gastroesophageal reflux disease, idiopathic pulmonary fibrosis, lung transplantation.

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive disorder of the lung characterized by fibrosis of the lower respiratory tract. The disease process is mediated by abnormal proliferation of mesenchymal cells and overproduction of collagen leading

to impaired gas exchange. Subepithelial fibrotic foci have been recognized as the histological hallmark of the disease,¹ but no other distinguishing markers of disease activity or prognosis have been identified via studies of serum, tissue, or bronchoalveolar lavage. There is agreement that certain risk factors are associated with IPF, which include tobacco use, environmental exposure, medications, infectious respiratory pathogens, and chronic aspiration.² Among the current theories on the pathogenesis of IPF is the contribution of gastroesophageal reflux disease (GERD) with chronic aspiration of gastric refluxate.

GERD is one of the most common gastrointestinal disorders seen by internists and gastroenterologists.¹ The well-known manifestations of reflux disease include heartburn, regurgitation, indigestion, and chest pain, whereas atypical signs and symptoms include laryngitis, sinusitis, cough, asthma, and recurrent pneumonias.^{2–4} Without treatment, GERD can cause esophageal complications of ulcerations, strictures, webs, Barrett esophagus, and cancer. In addition, untreated GERD may cause and exacerbate advanced lung diseases, including IPF, cystic fibrosis (CF), and chronic obstructive lung disease (COPD).

EVALUATION OF GASTROESOPHAGEAL REFLUX DISEASE

Although the common esophageal symptoms of GERD are generally easy to recognize and treat, the extra-esophageal symptoms, including cough, hoarse voice, asthma, and globus, are more difficult to associate with GERD. Over 30 years ago, 24-hour esophageal pH monitoring was developed to determine if patients had pathologic reflux and whether the reflux was temporally correlated with the symptom complex.⁵ Prior to this time, the presence of GERD was determined with a barium swallow. Radiologists were able to visualize reflux of barium from

Address Correspondence to:

Virginia R. Litle

Division of Thoracic and
Foregut Surgery
University of Rochester
Medical Center
Rochester, NY
Email: virginia_little@
urmc.rochester.edu

the stomach into the esophagus, especially when the patient was placed in the Trendelenberg position. This test, however, could not determine if a patient's symptoms were related to the visualized reflux. The advent of upper endoscopy, in which a fiber optic tube allows for direct visualization of the esophageal mucosa, permitted an assessment of esophageal inflammation. The endoscopic visualization of a hiatal hernia, Schatzki ring, or web is suggestive of reflux, whereas gross esophagitis or Barrett esophagus (intestinal metaplasia of the esophagus) gives direct evidence of gastric acid exposure within the esophagus. The mere presence of these conditions, however, does not imply a direct causal relationship to a symptom such as heartburn, regurgitation, chest pain, or extra-esophageal laryngeal or pulmonary manifestations.

The pathophysiology of GERD is multifactorial, and the role that each variable plays is unique to every patient. An incompetent lower esophageal sphincter (LES) or hiatal hernia correlates with reflux disease,⁶ and transient relaxation of the LES has been shown to occur in patients, accounting for reflux despite normal resting tone measured by esophageal manometry.⁷ The refluxate volume and the proximal extent (height) of the refluxed material will affect the contact time on the esophageal mucosa. These parameters have been shown to correlate with patient's symptoms, especially extra-esophageal complaints.^{8,9}

Acid Reflux

To standardize the definition of pathologic acid reflux, the DeMeester score was established years ago and is determined from the duration, frequency, and extent of distal esophageal acid exposure.^{10,11} Current technology available for measuring significant acid reflux by pH over a 24- to 48-hour period include a wire (catheter) system (Sandhill Scientific, Highlands Ranch, CO), a Slimline pH catheter system (Medtronic, Minneapolis, MN), and a wireless pH monitor (Bravo System, Medtronic). The catheter-based system is placed 5 cm above the manometrically located proximal border of the LES, whereas the wireless capsule (Bravo) is placed endoscopically 6 cm above the gastroesophageal junction. The systems record the esophageal pH every 4 to 6 seconds, and then the data are downloaded onto a computer for analysis. Numerous studies have been published that compare the advantages and limitations of each system.¹²

Non-Acid Reflux

An inherent limitation of pH studies is that acid reflux is defined when the pH of the refluxate is less than 4.0. Patients, however, may have symptoms of reflux with a pH of 5.0, and the standard pH testing would not detect this as a reflux event. In addition, a change in pH of 1 or 2 units (ie, from a pH of 6.5 to a pH of 4.5) may be just as important to the perception of heartburn as an absolute pH of 4.0. An improvement in the 24-hour pH test limitations has been multichannel intraluminal impedance (MII) testing (Sandhill Scientific), which examines reflux events independently of pH.^{12,13} A pH sensor is included within this specialized catheter, which allows for the measurements of both acid and non-acid reflux. MII can assess the composition (liquid, mixed, or gas) of the refluxate material and determine the proximal extent of the reflux. Proximal reflux has become an important parameter to examine because its presence has been associated with extra-esophageal reflux diseases such as cough, asthma, hoarse voice, laryngitis, and globus.¹⁴

II is able to detect reflux of fluid into the esophagus while patients are on standard proton pump inhibitor therapy. Hence, weakly acidic refluxate (above pH 4.0), which would go unnoticed in conventional pH measurements, can now be recorded and correlated with a patient's symptomatology. Recently, it has been shown that weakly acidic reflux may contribute to some of the extra-esophageal symptoms of reflux such as cough and globus.

Lastly, controversy in the literature exists as to what constitutes significant proximal reflux.¹⁵ Normative data are dependent on where the proximal probe is placed, that is, 15 or 20 cm above the LES. It is unclear whether the number of proximal events is the key parameter or the duration of the exposure is more important.¹⁶ To date, evidence for routine use of proximal pH monitoring is limited; however, further studies with newer technology may help to shed light on this aspect of reflux disease.

Medical treatment of GERD is generally aimed at decreasing acid production within the stomach. Proton pump inhibitors inhibit the proton pump within the parietal cell. H₂ blockers decrease acid production by blocking the H₂ receptor. Common over-the-counter antacids neutralize gastric acid and thus aid in decreasing the acidity of the refluxate. Despite neutralization of gastric acid, it should be remembered that reflux of gastric contents still occurs; just the pH of the refluxate has been altered.

If abnormal acid reflux is observed despite adequate medical treatment, then the patient should be offered surgical antireflux treatment with a

laparoscopic Nissen fundoplication or, if available, an endoscopic method to restore an LES.

Potentially overlooked factors that can contribute to GERD include delayed gastric emptying and esophageal dysmotility.^{17,18} In simple terms, gastroparesis can result in gastric contents refluxing back into the esophagus. Impaired esophageal motility can also worsen reflux, and no medications have been shown to be effective in increasing the percentage of peristaltic contractions or increasing LES tone. The test for esophageal dysmotility is esophageal manometry, which is typically done in conjunction with the 24-hour pH study or MII, and it provides measurements of esophageal contractility, sphincter pressures, and LES relaxation. A nuclear medicine gastric emptying study is obtained to assess delayed gastric emptying as a contributing factor to reflux.

Finally, decreased oropharyngeal secretions may decrease the neutralization of acidic gastric fluid, thereby increasing the apparent acid content time. Hypersensitivity of the esophageal mucosa may also account for GERD symptoms despite normal mucosa visualized on endoscopy or normal pH studies, and some investigators are studying the possibility of visceral neural pathway dysfunction.¹⁹

GASTROESOPHAGEAL REFLUX DISEASE, IDIOPATHIC PULMONARY FIBROSIS, AND OTHER ADVANCED LUNG DISEASES

The concept of GERD contributing to pulmonary pathology has been considered for several decades. The first studies of acid-mediated pulmonary damage were published in the 1950s when Teabeaut²⁰ showed that the degree of parenchymal damage became more severe with exposure to decreasing pH in a rabbit model. In the 1960s, dog experiments demonstrated the rapidity with which gastric acid is able to spread to the lung periphery and that aspiration of hydrochloric acid causes pneumonitis.^{21,22} Later, Glauser *et al.*²³ reconfirmed that aspiration of acid causes increased epithelial permeability and damage, hypoxemia, edema, and early inflammatory changes in the dog lung, and others reported significant pneumonitis and fibrotic proliferation in the rabbit model after injection of trisodium citrate and acid citrate.²⁴

GERD also has an established relationship with other pulmonary diseases such as asthma, CF, and COPD. A recent systematic review on the association between GERD and asthma showed the weighted average of GERD symptoms in asthma

patients to be 59.2% versus 38.1% in control patients (a 1.6-fold increase) as well as the average prevalence of abnormal esophageal pH, esophagitis, and hiatal hernia to be 50.9%, 37.3%, and 51.2%, respectively. These values are significantly higher than those of the general population (10%–20% for gastroesophageal reflux symptoms and 7%–16% for erosive esophagitis). Furthermore, the average prevalence of GERD as diagnosed by pH monitoring was 51%. Conversely, the prevalence of asthma in patients with GERD was 4.6% versus 3.9% for controls. Pooled odds ratios showed an overall ratio of 5.5 [95% confidence (CI): 1.9–15.8] for the prevalence of GERD in asthmatics and 2.3 (95% CI: 1.8–2.8) for the prevalence of asthma in patients with GERD.²⁵ In addition, patients with CF have increased GERD as measured by MII,²⁶ and it has been recommended that all patients with CF undergo a GERD workup,²⁷ although the long-term benefits of antireflux treatment for such patients is controversial.²⁸ Vega *et al.*²⁹ questioned 86 patients with COPD and found that the patients with GERD were twice as likely to have exacerbations of their lung disease. Not surprisingly, these exacerbations are associated with the increased costs of hospitalizations and emergency department and clinic visits. In pretransplant patients with COPD, Kempainen *et al.*³⁰ in Minnesota found a 57% incidence of reflux by 24-hour pH studies. Only 26% of these patients complained of heartburn or regurgitation.

Despite extensive animal model evidence of acid-mediated pulmonary damage and correlations with other pulmonary disorders, it was not until fairly recently that investigators began looking at GERD in relation to IPF. One of the first associations between pulmonary fibrosis and GERD was made in 1976 by Mays *et al.*,³¹ who looked at the frequency of hiatal hernia in patients with radiographic evidence of pulmonary fibrosis ($P > 0.001$). Further data were published by Raiha and colleagues,³² who evaluated 137 elderly patients referred for upper endoscopy in the setting of abdominal symptoms. Ninety-five of the patients underwent 24-hour pH monitoring and pulmonary evaluation, and the investigators concluded that the odds ratio for having respiratory symptoms was 8.7 (95% CI: 2.2–22.4) in subjects with a total reflux time of $>10\%$ versus those with a total reflux time of $>10\%$. Additionally, bilateral pleural adhesions and thickenings and bilateral pleural scars were seen significantly more often in those with regurgitation or large hiatal hernias.

In 1997, a large case-controlled study from a Veterans Administration hospital, which included more than 100,000 patients, provided evidence

for the association between erosive esophagitis and a number of respiratory diseases, including asthma, pneumonia, chronic bronchitis, and IPF. By multivariate logistic regression analysis, the odds ratio of erosive esophagitis being associated with pulmonary fibrosis was 1.36 (95% CI: 1.25–1.48).³³ Furthermore, Tobin *et al.*³⁴ compared 17 patients with biopsy-confirmed IPF with 8 controls with interstitial lung disease after ambulatory pH testing and found that 94% of the IPF patients had positive 24-hour pH studies versus only 50% of the other interstitial lung disease patients. There was more proximal esophageal acid exposure in the IPF patients, and overall, 75% of the patients with increased acid exposure had silent reflux (no symptoms).

This same concept was supported by Raghu *et al.*,³⁵ who looked at 65 patients with known IPF and 133 patients with intractable asthma. With esophageal manometry and 24-hour pH monitoring, they showed that 87% of patients with IPF had GERD, with the majority showing abnormal distal (76%) and proximal (63%) acid exposure. Abnormal esophageal acid exposure was also seen in 63% of patients receiving proton pump inhibitor therapy at the time of their pH testing, and this suggested that the IPF patient population may have a physiological resistance to standard doses of proton pump inhibitors. The prevalence of GERD was significantly lower in the intractable asthma population (68%, $P = 0.014$), and they were much less likely to have distal GERD ($P = 0.020$). Similarly to previously presented data, symptoms of GERD were present in only 47% of the IPF subjects. In addition, manometry testing (mean LES and percentage relaxation) was normal for all IPF patients in the study.

GASTROESOPHAGEAL REFLUX DISEASE AND LUNG TRANSPLANTATION

The lung transplant era has provided more evidence for the role of GERD in pulmonary disease. Bronchiolitis obliterans syndrome affects 50% to 60% of patients within 5 years after transplant. Theories on the mechanism of this devastating complication include acute rejection, ischemia reperfusion injury, cytomegalovirus infection, perioperative factors, and GERD. Gastrointestinal complaints are very common post-transplant, with potential causes including intraoperative vagal trauma and postoperative immunosuppressive therapy.^{36,37} Palmer and colleagues³⁸ studied the prevalence of GERD in a cohort of

23 patients both pre-transplant and post-transplant (24-hour pH monitoring, motility, and gastric emptying studies) and found that post-transplant, the total acid contact time increased a mean of 3.7% ($P = 0.03$), and the supine acid contact time increased a mean of 6.4% ($P = 0.019$). Overall, 35% of the patients had abnormal acid contact time pre-transplant, and 65% had it post-transplant. No correlation between acid contact time and motility or gastric emptying was established. Finally, confirming previous data, they found that only 20% of the patients with abnormal GERD were symptomatic. Bronchiolitis obliterans syndrome, synonymous with chronic lung rejection, is diagnosed by a >20% decline in pulmonary function tests from baseline. Eighteen post-lung transplant patients with documented GERD underwent antireflux surgery, and 67% of the patients showed improvement in their forced expiratory volume in 1 second. In addition, there were no perioperative deaths.³⁹

Several groups have looked at the prevalence of GERD in patients with end-stage lung disease awaiting transplantation. D'Ovidio and colleagues⁴⁰ studied 78 patients with various advanced lung diseases, including IPF (26 patients), scleroderma, COPD, and CF, and they found that 72% had a hypotensive LES, 33% had dysmotility of the esophageal body, 44% showed prolonged gastric emptying, and 38% had abnormal 24-hour pH testing. In 109 patients awaiting transplantation, reflux symptoms were not predictive of the presence of reflux (67% sensitive), a hypotensive LES was very prevalent (55%), and a significant percentage of those with reflux had impaired esophageal contractility (47%).⁴¹ In an IPF-only study ($n = 30$), 67% were found to have abnormal esophageal reflux, and among these, 65% had a hypotensive LES. Abnormal esophageal motility was also more common in patients with GERD ($P = 0.03$), and typical reflux symptoms were not a reliable screening test (sensitivity: 65%, specificity: 71%).⁴²

Up to two-thirds of patients with end-stage IPF have GERD by 24-hour pH monitoring, and typical reflux symptoms are insufficient to distinguish who refluxes.⁴³ Although in a small series of 4 patients, medical treatment for GERD was considered for IPF patients,⁴⁴ several larger groups have advocated antireflux surgery with a laparoscopic Nissen fundoplication to control GERD in patients with end-stage lung disease and other extra-esophageal manifestations.^{43,45,46} In a series of 74 patients with GERD and extra-esophageal symptoms (including bronchospasm) who underwent laparoscopic Nissen fundoplication, there was no perioperative mortality, there was 5% morbidity, and all symptoms except

laryngitis improved significantly.¹⁵ Patti *et al.*¹⁵ recommended a Nissen fundoplication before or shortly after lung transplantation for IPF. Bueno's group¹⁶ performed laparoscopic Nissen fundoplication on 14 pretransplant patients with IPF and found a stabilization of oxygen requirements in comparison with IPF patients who did not have surgical GERD treatment.

CONCLUSION

GERD is more prevalent in patients with IPF than is commonly acknowledged. Patients with IPF and other advanced lung diseases should be referred to an esophageal diagnostic laboratory for pH monitoring, esophageal manometry, and impedance testing. The growing body of literature supports an association between GERD and advanced lung disease, but increased use of esophageal diagnostics and lung biopsies may help prove a causal relationship and allow for less invasive treatment options than lung transplantation for end-stage lung disease.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Holtmann G. Reflux disease: the disorder of the third millennium. *Eur J Gastroenterol Hepatol* 2001; 13(suppl 1): S5–S11.
- DeVault KR. Overview of therapy for the extraesophageal manifestations of gastroesophageal reflux disease. *Am J Gastroenterol* 2000; 95(suppl): S39–S44.
- Vakil N, Veldhuyzen van Zante S, Kahrilas P, et al. The Montreal definition and classification of gastroesophageal reflux disease (GERD): a global evidenced-based consensus. *Am J Gastroenterol* 2006; 1001: 1900–1920.
- Champion GL, Richter JE. Atypical presentation of gastroesophageal reflux disease: chest pain, pulmonary, and ear, nose, throat manifestations. *Gastroenterologist* 1993; 1: 18–33.
- DeMeester TR, Wang CI, Wernly JA, et al. Technique, indications, and clinical use of 24 hour esophageal pH monitoring. *J Thorac Cardiovasc Surg* 1980; 79: 656–670.
- Fein M, Ritter MP, DeMeester TR, et al. Role of the lower esophageal sphincter and hiatal hernia in the pathogenesis of gastroesophageal reflux disease. *J Gastrointestinal Surg* 1999; 3: 405–410.
- Iwakiri K, Hayange Y, Kotohira M, et al. Pattern of acid reflux during transient lower esophageal sphincter relaxation (TLESRs) differ among patients with mild and severe reflux esophagitis and healthy subjects. *Gastroenterology* 2004; 126(suppl 2): M2007.
- Sifrim D. Relevance of volume and proximal extent of reflux in gastroesophageal reflux disease. *Gut* 2005; 54: 175–178.
- Fass R, Achem SR, Harding S, et al. Review article: supra-oesophageal reflux disease and the role of night-time gastro-oesophageal reflux. *Aliment Pharmacol Ther* 2004; 20(suppl 9): 26–38.
- Johnson LF, DeMeester TR. Twenty-four hour pH monitoring of the distal esophagus. A quantitative measure of gastroesophageal reflux. *Am J Gastroenterol* 1974; 62: 325–332.
- Johnson LF, DeMeester TR. Development of the 24-hour intraesophageal pH monitoring system composite scoring system. *J Clin Gastroenterol* 1986; 8(suppl 1): 52–58.
- Hirano I, Richter JE. ACG practice guidelines: esophageal reflux testing. *Am J Gastroenterol* 2007; 102: 668–685.
- Mainie I, Tutuian R, Shay S, et al. Acid and non-acid reflux in patients with persistent symptoms despite acid suppressive therapy: a multicenter study using combined ambulatory impedance-pH monitoring. *Gut* 2006; 55: 1398–1402.
- Anandasabapathy S, Jaffin BW. Multichannel intraluminal impedance in the evaluation of patients with persistent globus on proton pump inhibitor therapy. *Ann Otol Rhinol Laryngol* 2006; 115: 563–570.
- Cool M, Poelmans J, Feenstra L, et al. Characteristics and clinical relevance of proximal esophageal pH monitoring. *Am J Gastroenterol* 2004; 99: 2317–2323.
- Sharma N, Agrawal A, Freeman J, et al. An analysis of persistent symptoms in acid-suppressed patients undergoing impedance-pH monitoring. *Clin Gastroenterol Hepatol* 2008; 6: 521–524.
- Buckles DC, Sarosick I, MacMillin C, et al. Delayed gastric emptying in gastroesophageal reflux disease: reassessment with new methods and somatic correlations. *Am J Med Sci* 2004; 327: 1–4.
- Shay SS, Egli D, McDonald C, et al. Gastric emptying of solid food in patients with gastroesophageal reflux. *Gastroenterology* 1987; 92: 459–465.
- Rodriguez-Stanley S, Robinson M, Easnest DL, et al. Esophageal hypersensitivity may be a major cause of heartburn. *Am J Gastroenterol* 1999; 94: 628–631.
- Teabeaut JR. Aspiration of gastric contents: an experimental study. *Am J Pathol* 1952; 28: 51–62.
- Hamelberg W, Bosomworth PP. *Anesth Analg Curr Res* 1964; 43: 669–677.
- Greenfield LJ, Singleton RP, McCaffree DR, et al. Pulmonary effects of experimental graded aspiration of hydrochloric acid. *Ann Surg* 1969; 170: 74–86.
- Glauser FL, Millen JE, Falls R. Effects of acid aspiration on pulmonary alveolar epithelial membrane permeability. *Chest* 1979; 76: 201–205.
- Mitsuhashi T, Shimazaki M, Chanoki Y, et al. Experimental pulmonary fibrosis induced by trisodium citrate and acid-citrate-dextrose. *Exp Mol Pathol* 1985; 42: 261–270.
- Havemann BD, Henderson CA, El-Serag HB. The association between gastro-oesophageal reflux disease

- and asthma: a systematic review. *Gut* 2007; 56: 1654–1664.
26. Blondeau K, Dupont L, Mertens V, et al. Gastroesophageal reflux and aspiration of gastric contents in adult patients with cystic fibrosis. *Gut* 2008; 57: 1049–1055.
27. Brodzicki J, Trawińska-Bartnicka M, Korzon M. Frequency, consequences and pharmacological treatment of gastroesophageal reflux in children with cystic fibrosis. *Med Sci Monit* 2002; 8: CR529–37.
28. Boesch RP, Acton JD. Outcomes of fundoplication in children with cystic fibrosis. *J Pediatr Surg* 2007; 42: 1341–1344.
29. Rascon-Aguilar IE, Pamer M, Wludyka P, et al. Role of gastroesophageal reflux symptoms in exacerbations of COPD. *Chest* 2006; 130: 1096–1101.
30. Kempainen RR, Savik K, Whelan T, et al. High prevalence of proximal and distal gastroesophageal reflux disease in advanced COPD. *Chest* 2007; 131: 1666–1671.
31. Mays EE, Dubois JJ, Hamilton GB. Pulmonary fibrosis associated with tracheobronchial aspiration: a study of the frequency of hiatal hernia and gastroesophageal reflux in interstitial fibrosis of obscure etiology. *Chest* 1976; 69: 512–515.
32. Raiha I, Manner R, Hietanen E. Radiographic pulmonary changes of gastro-oesophageal reflux disease in elderly patients. *Age Ageing* 1992; 21: 250–255.
33. el-Serag HB, Sonnenberg A. Comorbid occurrence of laryngeal or pulmonary disease with esophagitis in United States military veterans. *Gastroenterology* 1997; 113: 755–760.
34. Tobin RW, Pope CE, Pellegrini CA. Increased prevalence of gastroesophageal reflux in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1988; 158: 1804–1808.
35. Raghu G, Freudenberger TD, Yang S, et al. High prevalence of abnormal acid gastro-oesophageal reflux in idiopathic pulmonary fibrosis. *Eur Respir J* 2006; 27: 136–142.
36. Sodhi SS, Guo JP, Maurer AH, et al. Gastroparesis after combined heart and lung transplantation. *J Clin Gastroenterol* 2002; 34: 34–39.
37. Maes BD, Vanwalleghe J, Kuypers D, et al. Differences in gastric motor activity in renal transplant recipients treated with FK-506 versus cyclosporine. *Transplantation* 1999; 68: 1482–1485.
38. Young LR, Hadjiliadis D, Davis D, et al. Lung transplantation exacerbates gastroesophageal reflux disease. *Chest* 2003; 124: 1689–1693.
39. Lau CL, Palmer SM, Howell DN, et al. Laparoscopic antireflux surgery in the lung transplant population. *Surg Endosc* 2002; 16: 1674–1678.
40. D'Ovidio F, Singer LG, Hadjiliadis D, et al. Prevalence of gastroesophageal reflux in end-stage lung disease candidates for lung transplant. *Ann Thoracic Surg* 2005; 80: 1254–1261.
41. Sweet MP, Herbella FAM, Leard L, et al. The prevalence of distal and proximal gastroesophageal reflux in patients awaiting lung transplantation. *Ann Surg* 2006; 244: 491–497.
42. Sweet MP, Patti MG, Leard LE, et al. Gastroesophageal reflux in patients with idiopathic pulmonary fibrosis referred for lung transplantation. *J Thorac Cardiovasc Surg* 2007; 133: 1078–1084.
43. Patti MG, Tedesco P, Golden J, et al. Idiopathic pulmonary fibrosis: how often is it really idiopathic? *J Gastrointest Surg* 2005; 9: 1053–1056.
44. Raghu G, Yang ST, Spada C, et al. Sole treatment of acid gastroesophageal reflux in idiopathic pulmonary fibrosis: a case series. *Chest* 2006; 129: 794–800.
45. Fernando HC, El-Sherif A, Landreneau RJ, et al. Efficacy of laparoscopic fundoplication in controlling pulmonary symptoms associated with gastroesophageal reflux disease. *Surgery* 2005; 138: 612–616.
46. Linden PA, Gilbert RJ, Yeap BY, et al. Laparoscopic fundoplication in patients with end-stage lung disease awaiting transplantation. *J Thorac Cardiovasc Surg* 2006; 131: 438–446.

“End-Stage” Pulmonary Fibrosis in Sarcoidosis

Alvin T. Teirstein, MD, and Adam S. Morgenthau, MD

Mount Sinai School of Medicine, New York, NY

ABSTRACT

Pulmonary fibrosis is an unusual “end stage” in patients with sarcoidosis. Fibrosis occurs in a minority of patients, and presents with a unique physiologic combination of airways dysfunction (obstruction) superimposed on the more common restrictive dysfunction. Imaging techniques are essential to the diagnosis, assessment and treatment of pulmonary fibrosis. Standard chest radiographs and CT scans may reveal streaks, bullae, cephalad retraction of the hilar areas, deviation of the trachea and tented diaphragm. Positive gallium and PET scans indicate residual reversible granulomatous disease and are important guides to therapy decisions. Treatment, usually with corticosteroids, is effective in those patients with positive scans, but fibrosis does not improve with any treatment. With severe functional impairment and patient disability, pulmonary hypertension and right heart failure may supervene for which the patient will require treatment. Oxygen, careful diuresis, sildenafil and bosentan may be salutary. These patients are candidates for lung transplantation. *Mt Sinai J Med* 76:30–36, 2009. © 2009 Mount Sinai School of Medicine

Key Words: Fibrosis, airways obstruction, bullae, tented diaphragm, hilar retraction, gallium and PET scans, pulmonary hypertension, right heart failure.

There is no staging system for sarcoidosis; there is only a system for describing chest radiological patterns. Although Siltzbach and Scadding's original 3 radiographic stages imply an orderly chronology, stage I (hilar adenopathy only) may be chronic

and persist forever, whereas stages II and III (parenchymal disease) may occur acutely.^{1,2} Because of the inadequacy of this staging and its frequent misinterpretation, radiographs exhibiting fibrotic changes have been added to the original 3 patterns and labeled stage IV. In contrast, these patients almost always have long-lasting chronic disease and are the most difficult to treat. They encompass those patients who are least responsive to treatment, are most likely to have severe pulmonary dysfunction, and have the highest morbidity and mortality.

There are few reports on the frequency of pulmonary fibrosis in sarcoidosis. An estimate of the prevalence of fibrotic pulmonary sarcoidosis can be made from studies that suggest 10% to 20% of patients with sarcoidosis present with stage IV chest radiographs.³ In ACCESS (A Case Controlled Etiology Study of Sarcoidosis), a stage IV chest radiograph was found in 5.4% of sarcoidosis cases.⁴

Sarcoidosis usually declares its ultimate manifestations during the first 2 years after diagnosis.⁵ Fortunately, most patients improve spontaneously or stabilize with little or no therapy. A subset will continue to smolder with pulmonary, ocular, and cutaneous lesions that wax and wane. Unfortunately, a minority of patients will progress to fibrosis, which leaves fibrotic lungs, eyes, and/or skin. Why some patients progress to fibrosis has bedeviled investigators for generations. In the hope of offering early therapy, researchers have proposed numerous clinical and biochemical markers that predict a fibrotic course. To date, there is no medication proven to be capable of reversing fibrosis, but there is hope that treatment can arrest fibrosis of reversible granulomas that persist among the fibrotic elements. Techniques that are purported to differentiate fibrotic tissue from granulomatous tissue have grown in importance.

Address Correspondence to:

Alvin T. Teirstein

Mount Sinai School of Medicine
New York, NY

Email: alvin.teirstein@mountsinai.org

PATHOLOGY

Usually, pulmonary sarcoidosis is marked by diffuse infiltration of the lung parenchyma by monotonous,

well-formed nonnecrotizing granulomas most prominent along the bronchovascular bundles. The airways are usually spared.⁶ In those patients whose disease progresses to fibrosis, however, manifestations of airway involvement abound. Bronchiectatic microcysts and honeycombing, formation of large bullae, cephalad retraction of the hilar structures with distortions of the tracheal pathway, tenting of the diaphragm, and pneumothorax may be seen.⁷ The residua of repeated infections, including atelectasis, pleural thickening, and aspergillomas, may compound this chaotic disruption of lung tissue (Figures 1–3). Distortion of the airways has been thought to be due to obstruction of small airways and/or fibrotic retraction of the parenchyma. However, recent studies have implicated a variety of genetic abnormalities and fibrogenic cytokines that are increased in sarcoidosis patients and lead to irreversible pulmonary fibrosis.

GENETICS

Genetic factors are instrumental in fibrosis. Recent evidence suggests that there is reduced expression of prostaglandin-endoperoxide synthase 2 (PTGS2) or cyclooxygenase-2 in fibrotic lung diseases, including sarcoidosis. Reduced expression of PTGS2 removes an important inhibitory control that may limit progressive fibrotic lung disease. PTGS2 is involved in the enzymatic conversion of arachidonic acid to prostanoids. Prostanoids provide an important homeostatic control in normal tissues and regulate inflammation. Prostaglandin E₂ (PGE₂) is the predominant prostanoid produced in the normal lung. It is also the main prostanoid produced by lung fibroblasts, and it exerts fibroprotective effects by inhibiting their proliferation, differentiation into myofibroblasts, and collagen synthesis. Several studies have demonstrated that PGE₂ levels in the lung are reduced in pulmonary fibrosis patients by approximately 50%, despite increased levels of proinflammatory and profibrotic cytokines such as interleukin-1 (IL-1), tumor necrosis factor- α , and transforming growth factor- β (TGF- β). Additional studies have shown that reduced levels of PGE₂ result from decreased PTGS2 expression. Reduced PTGS2 expression, in turn, results from polymorphisms within the gene.⁸

According to Hill and colleagues,⁹ carriage of the -765C allele of the -765G>C polymorphism confers susceptibility to sarcoidosis and predicts poor outcomes. In their study, 47% of the subjects with the -765C allele demonstrated persistent disease versus 18% of the healthy subjects and 23% of the subjects with nonpersistent sarcoidosis. Patients

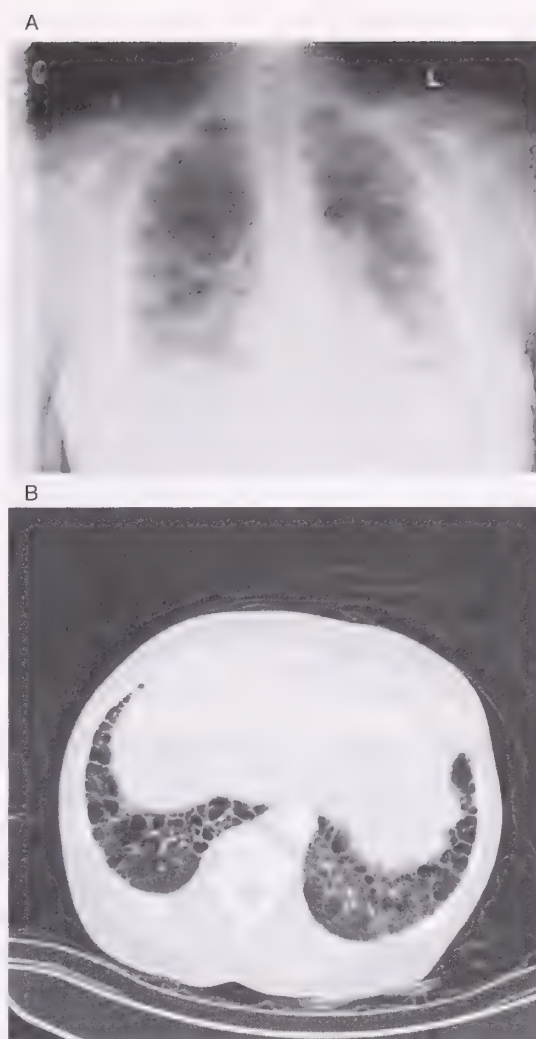


Fig 1. (A) Standard chest PA radiograph and (B) computed tomography scan of a 46-year-old woman with biopsy-proven sarcoidosis showing large upper and lower zone bullae.

with a stage 0 or I radiograph were defined as having nonpersistent disease. Patients with stage IV radiographs were defined as having persistent disease. Patients with stage II or III radiographs, were defined as having persistent disease if they also had pulmonary sarcoidosis for more than 2 years or were treated for their pulmonary disease and demonstrated abnormal pulmonary function [forced vital capacity (FVC), total lung capacity, and/or diffusing capacity of the lung for carbon monoxide <80% of the predicted values]. The authors concluded that the -765G>C polymorphism could potentially be used as a prognostic marker for poor outcomes in this population of sarcoidosis patients.

Polymorphisms in the genes encoding for the 3 isoforms of TGF- β have been identified and linked to fibrosis. TGF- β 1 appears to play a

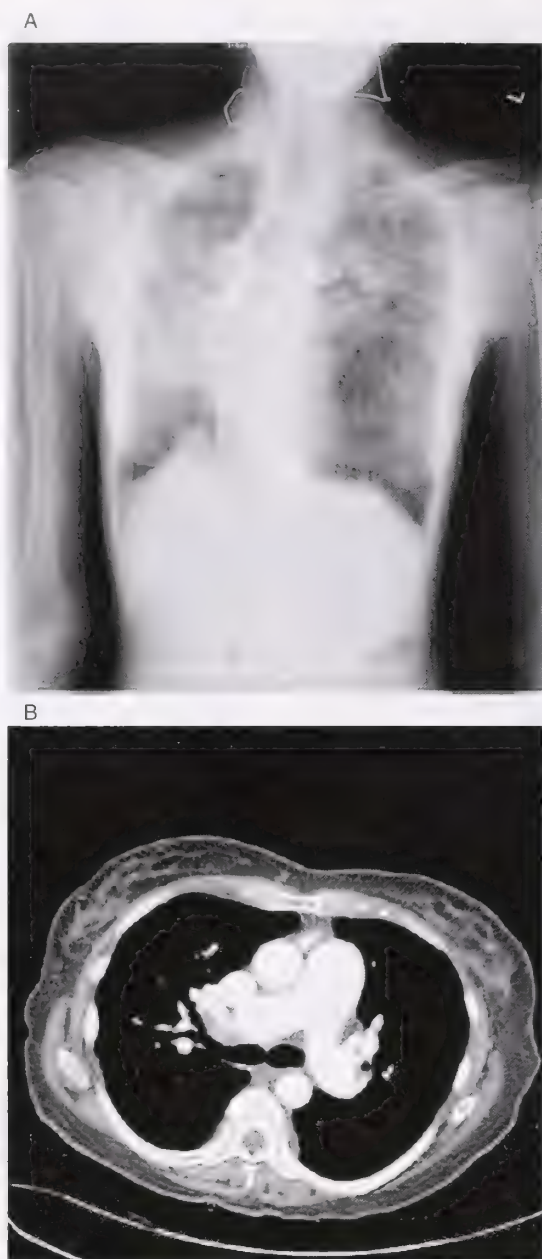


Fig 2. (A) Standard chest PA radiograph of a 66-year-old woman with biopsy-proven sarcoidosis exhibiting stage IV radiographic features: retraction of the hilar areas, cephalad streaks, deviation of the trachea to the right, tenting of the diaphragm, and nodules. (B) Computed tomography scan of the same patient with contrast showing a large pulmonary outflow tract indicating the presence of pulmonary hypertension.

central role in idiopathic pulmonary fibrosis and cystic fibrosis. Recent studies have suggested that TGF- β 2 and TGF- β 3 are associated with cutaneous fibrosis in patients with systemic scleroderma. Kiuit and colleagues¹⁰ examined whether TGF- β polymorphisms could be found at higher rates in 145 patients with sarcoidosis compared to 315



Fig 3. Computed tomography scan of a 72-year-old woman with biopsy-proven sarcoidosis exhibiting honeycombing of the lower lobes.

healthy, unrelated control subjects. Interestingly, TGF- β 3 polymorphisms were more frequent in fibrotic patients than in patients with acute/self-remitting disease ($P < 0.05$, odds ratio = 5.1).

Jonth and associates¹¹ examined DNA from participants in ACCESS to determine whether genetic variants in the TGF- β 1 gene were found at higher rates in sarcoidosis patients versus controls and were associated with indicators of disease severity. They concluded that a haplotype of the TGF- β 1 gene (-509C and codon 10T) resulted in reduced levels of the TGF- β 1 protein, which was associated with severe granulomatous disease in the sarcoidosis patients. Although the exact mechanism for the deficiency was not elucidated, the study identified a potential marker of disease severity in sarcoidosis patients.

An oxidant-antioxidant imbalance has been implicated in pulmonary fibrosis. Markart *et al.*¹² reported the presence of increased antioxidants in the bronchoalveolar lavage fluid (BALF) in the bleomycin animal model of pulmonary fibrosis and in the lungs of patients with idiopathic pulmonary fibrosis, chronic sarcoidosis, and hypersensitivity pneumonitis. They postulated that these antioxidants were not capable of correcting the oxidative imbalance.

IMMUNOPATHOGENESIS

There has been limited progress in the development of efficacious therapies for chronic fibrotic lung disease. Experimental models have demonstrated that

cytokine networks are operative in fibrosis. Cytokine profiles in immune-mediated granulomatous inflammation may be dominated by T helper 1 (Th1) cytokines [interferon- γ (IFN- γ)], IL-2, or T helper 2 (Th2) cytokines (IL-4, IL-5, and IL-13). The cytokines are modulated in response to an inciting agent or antigen. These cytokines have pleomorphic effects, but it has long been recognized that IFN- γ suppresses fibroblast proliferation and collagen deposition.¹³

There is robust evidence that granulomatous inflammation in sarcoidosis involves a highly polarized Th1 immune response. In experimental models, granulomatous inflammation is down-regulated when there is clearance of the antigen. Failed clearance of the antigen results in persistent antigenic stimulation, which signals the host to modify its defenses and to enter a transition phase, which is characterized by a hybrid Th1/Th2 phenotype or an isolated Th2 cytokine phenotype. The elaboration of type 2 cytokines aids in attacking and clearing the inciting antigen. Type 2 cytokines eventually activate resident fibroblasts to proliferate and deposit collagen, causing chronic lung inflammation and fibrosis. Fibroblastic proliferation and collagen deposition are the host's protective barricade against the inciting antigen.¹⁴

There are few data on cytokine profiles in late stages of fibrotic sarcoidosis for an assessment of the contribution of Th1 or Th2 cytokines to the fibrotic process. It is conceivable, however, that patients with fibrotic sarcoidosis have genetically endowed immune systems that switch to a profibrotic Th2 response later in the disease either to down-regulate the proinflammatory Th1 response or to clear the inciting antigen.

ANGIOGENESIS

Recent evidence suggests that certain cytokines and chemokines are involved in the pathogenesis of angiogenesis. Angiogenesis is the process of new blood vessel growth, which is critical for wound healing and may contribute to the fibroproliferation and extracellular matrix deposition observed in idiopathic pulmonary fibrosis and fibrotic sarcoidosis. IFN- γ is one of several Th1-type cytokines that orchestrate granuloma formation. Recent studies suggest that IFN- γ has pleotrophic properties, one of which is to regulate the CXC chemokines. The CXC chemokines, in turn, modulate angiogenesis.¹⁵

Agostini *et al.*¹⁶ demonstrated that there is enhanced expression of interferon-inducible protein 10 (IP-10; a CXC chemokine) in sarcoid tissues and that increased BALF IP-10 levels are associated with

the degree of T cell alveolitis in sarcoidosis patients. The authors postulated that T cells within the IP-10 milieu migrate to sites of persistent inflammation in pulmonary sarcoidosis. Katoh and associates¹⁷ reported elevated BALF concentrations of IP-10 and monokine induced by IFN- γ (another CXC cytokine) in patients with sarcoidosis and eosinophilic pneumonia. Finally, Sekiya and colleagues¹⁸ demonstrated a strong correlation of elevated serum levels of vascular endothelial growth factor (a growth factor in angiogenesis) with clinical parameters of disease severity in sarcoidosis. Taken together, these studies suggest that IFN- γ -induced CXC chemokines modulate angiogenesis in chronic fibrotic lung disease.

The fibrosis that results from angiogenesis and possibly the process of angiogenesis itself lead to destruction of the capillary bed, hypoxemia, and sometimes pulmonary hypertension. Less than 5% of all sarcoidosis patients develop pulmonary hypertension, but it is relatively common in advanced disease. Although the majority of sarcoidosis patients with pulmonary hypertension have a stage IV chest radiograph, the severity of pulmonary hypertension does not correlate well with the degree of pulmonary fibrosis and hypoxemia. There are mechanisms, in addition to those involved in pulmonary fibrosis, that contribute to the development of sarcoidosis-associated pulmonary hypertension. These include extrinsic compression of large pulmonary arteries by mediastinal or hilar lymph nodes and granulomatous invasion of the pulmonary vasculature that may result in pulmonary veno-occlusive disease and pulmonary vasoconstriction. Although sarcoidosis-associated pulmonary hypertension was initially considered a rare complication of chronic disease, it is now recognized as an important complication that can result in significant morbidity and mortality.

SYMPTOMS AND SIGNS

Dyspnea and cough may occur in all phases of sarcoidosis. These symptoms do not differentiate between those patients with diffuse parenchymal granulomatous infiltrations and those with fibrosis. However, cough associated with wheezes and rhonchi and the presence of crackles in patients with dyspnea who are resistant to therapy strongly suggest the presence of fibrosis.

PULMONARY FUNCTION

Not surprisingly, with the inception of fibrosis, studies of airway dysfunction and air trapping

are superimposed on the more common finding of restrictive dysfunction. Decreased flow rates, decreased $FEV_1/FVC \times 100$ ratios (FEV_1 is the forced expiratory volume in 1 second), and increased residual volume compound the usual restricted pattern characterized by decreased FEV_1 and lung volumes with a loss of gas exchange capability, as reflected in decreased single-breath diffusing capacity of the lung for carbon monoxide. These findings are so distinctive that decades ago, Miller *et al.*¹⁹ proposed a physiological staging system for sarcoidosis based on the various pulmonary function aberrations.

IMAGING

In general, radiological alterations parallel the pathological abnormalities. Standard chest radiographs reveal distortion of the lung and mediastinal anatomy by thick fibrotic bands, bullous transformation, and segmental and even lobular atelectasis (usually the right middle lobe; Figure 2). Computed tomography (CT) scans are much more revealing.²⁰ However, for clinical purposes, CT scans are best used in patients presenting with hemoptysis in a search for aspergillomas, which may occupy a bullous structure (Figure 3), or when the standard chest radiograph is atypical for sarcoidosis. Because fibrotic pulmonary sarcoidosis is a chronic disease that requires radiographic and physiological monitoring, standard chest radiographs are usually sufficient to follow these relatively young patients, and the excessive radiation and expense of CT scans are avoided.²¹ Recent studies of inspiratory and expiratory CT scans suggest that some degree of air trapping is present in most patients with pulmonary sarcoidosis, even those without pulmonary fibrosis.²² Although the correlation of imaging studies and pulmonary function is generally good, there are a large number of patients with discordance. The correlation is usually quite close between CT findings and pulmonary dysfunction in symptomatic patients with pulmonary fibrosis. Almost all of these patients manifest CT and standard chest imaging suggesting extensive fibrosis accompanied by some level of restriction and obstruction and decreased single-breath diffusing capacity of the lung for carbon monoxide.²³

TREATMENT

There are few, if any, valid randomized studies of therapy in sarcoidosis. Although minor improvements in chest radiographs and pulmonary function

have been reported, Cochrane reviews evaluating corticosteroid and other immunosuppressants in the treatment of sarcoidosis question the efficacy of these treatments.²⁴ Thomas and Hunninghake²⁵ recommended observation without treatment in most patients with sarcoidosis because the disease has a well-known propensity to spontaneously wax and wane, and many patients improve spontaneously. When granulomatous pathology is major and pulmonary function is significantly impaired, oral corticosteroid therapy is usually effective. However, fibrosis does not improve and may progress relentlessly. There is little evidence to support corticosteroid or immunosuppressive treatment of fibrotic lung disease unless the presence of active granulomas can be demonstrated. Thus, multiple indices of activity in sarcoidosis have emerged. Clinical indices include active skin lesions, uveitis, hypercalcemia, enlarged lacrimal and parotid glands, and enlarged liver and spleen.²⁶ A long list of biochemical studies of activity have been proposed, beginning with serum lysozyme and angiotensin converting enzyme levels, which may indicate active granulomatous disease.²⁷ Unfortunately, these clinical and serologic indices may indicate sarcoidosis activity but do not point to the loci of activity and have little clinical utility in patients with pulmonary fibrosis. Bronchoalveolar lavage lymphocytosis was heralded as a sensitive index of activity more than 20 years ago,²⁸ but it has little practical application in patient care. Imaging techniques appear to have the most utility when the clinician is faced with the challenge of treating or discontinuing treatment in a symptomatic patient with evidence of pulmonary fibrosis. Lynch²⁹ extensively reviewed those CT findings that indicate fibrotic, non-reversible lesions. However, identifying the presence of residual active, reversible lesions in a background of fibrosis is much more difficult. Gallium scans have been valuable in detecting active granulomas, and 2-[¹⁸F]-fluoro-2-deoxy-d-glucose positron emission tomography (PET) scans are proving to be even more sensitive.³⁰ When a patient presents with a stage IV chest radiograph or CT suggesting pulmonary fibrosis, a gallium or PET scan showing pickup of the radioisotope in the pulmonary parenchyma can serve as a guide for treatment. Conversely, in the case of a patient with pulmonary fibrosis receiving therapy, negative nuclear scans allow the reduction or discontinuation of medications.

Which therapy is indicated? Most patients with sarcoidosis, even with active parenchymal granulomas, are asymptomatic with acceptable pulmonary function and require no treatment. In patients with

pulmonary fibrosis and disabling pulmonary dysfunction, treatment is usually indicated for the symptomatic patient who has a positive gallium scan or PET scan with uptake of the radionuclide in the lungs. In these patients, systemic corticosteroid therapy, usually prednisone (30–40 mg/day), is indicated. Although many clinicians add a steroid-sparing agent such as hydroxychloroquine, methotrexate, or aziothioprine, there are no controlled trials that have proven the efficacy of these drugs.²² Similarly, reports of improvement with tumor necrosis factor blocking agents such as infliximab and etanercept remain anecdotal. Finally, in patients with severe dyspnea, hypoxemia, and pulmonary hypertension, supportive therapy with oxygen and treatment of right heart failure are palliative and important. Sildenafil and bosentan, agents directed at lowering pulmonary artery pressure, are often prescribed.^{31,32} These latter patients should be evaluated for lung transplantation. Because many die awaiting a donor organ, they should be placed on the waiting list as soon as irreversible pulmonary fibrosis is demonstrated.³³

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Siltzbach LE. Sarcoidosis: clinical features and management. *Med Clin North Am* 1967; 51: 483–502.
2. Scadding JG. Prognosis of intrathoracic sarcoidosis in England: a review of 136 cases after five years' observation. *Br Med J* 1961; 2: 1165–1172.
3. Nunes H, Brillet P, Valeyre D, et al. Imaging in sarcoidosis. *Semin Respir Crit Care Med* 2007; 28: 102–120.
4. Baughman RP, Teirstein AS, Judson MA, et al., for the ACCESS Research Group. Clinical characteristics of patients in a case control study of sarcoidosis. *Am J Respir Crit Care Med* 2001; 164: 1885–1889.
5. Judson MA, Baughman RP, Thompson BW, et al., for the ACCESS Research Group. Two year prognosis of sarcoidosis: the ACCESS experience. *Sarcoidosis Vasc Diffuse Lung Dis* 2003; 20: 204–211.
6. Longcope WT, Freiman DG. A study of sarcoidosis: based on a combined investigation of 160 cases including 30 autopsies from the Johns Hopkins Hospital and Massachusetts General Hospital. *Medicine* 1952; 31: 1–132.
7. Rosen Y, Vuletin JC, Pertschuk LP, Silverstein E. Sarcoidosis: from the pathologist's vantage point. *Pathol Annu* 1979; 14: 405–439.
8. Petkova DK, Clelland CA, Ronan JE, et al. Reduced expression of cyclooxygenase (COX) in idiopathic pulmonary fibrosis and sarcoidosis. *Histopathology* 2003; 43: 381–386.
9. Hill MR, Papafili A, Booth H, et al. Functional prostaglandin-endoperoxide synthase 2 polymorphism predicts poor outcome in sarcoidosis. *Am J Respir Crit Care Med* 2006; 174: 915–922.
10. Kruit A, Grutters JC, Ruven HJ, et al. Transforming growth factor-beta gene polymorphisms in sarcoidosis patients with and without fibrosis. *Chest* 2006; 129: 1584–1591.
11. Jonth AC, Silveira L, Fingerlin TE, et al., for the ACCESS Group. TGF-beta 1 variants in chronic beryllium disease and sarcoidosis. *J Immunol* 2007; 179: 4255–4262.
12. Markart P, Luboeinski T, Korfei M, et al. Alveolar oxidative stress is associated with elevated levels of non-enzymatic low-molecular-weight antioxidants in patients with different forms of chronic fibrosing interstitial lung diseases. *Antioxid Redox Signal* 2008; doi://10.1089/ARS.2008.2105. Available at: <http://www.liebertonline.com>.
13. Kunkel SL, Chensue SW, Lukacs N, Hogaboam C. Cytokine phenotypes serve as a paradigms for experimental immune-mediated lung diseases and remodeling. *Am J Respir Cell Mol Biol* 2003; 29: S63–S66.
14. Moller DR. Pulmonary fibrosis of sarcoidosis: new approaches, old ideas. *Am J Respir Cell Mol Biol* 2003; 29: S37–S41.
15. Sugiyama K, Mukae H, Ishii H, et al. Elevated levels of interferon gamma-inducible protein-10 and epithelial neutrophil-activating peptide-78 in patients with pulmonary sarcoidosis. *Respirology* 2006; 11: 708–714.
16. Agostini C, Cabrelle A, Calabrese F, et al. Role for CXCR6 and its ligand CXCL16 in the pathogenesis of T-cell alveolitis in sarcoidosis. *Am J Respir Crit Care Med* 2005; 172: 1290–1298.
17. Katoh S, Fukushima K, Matsumoto N, et al. Accumulation of CXCR3-expressing eosinophils and increased concentration of its ligands (IP10 and Mig) in bronchoalveolar lavage fluid of patients with chronic eosinophilic pneumonia. *Int Arch Allergy Immunol* 2005; 137: 229–235.
18. Sekiya M, Ohwada A, Miura K, et al. Serum vascular endothelial growth factor as a possible prognostic indicator in sarcoidosis. *Lung* 2003; 181: 259–265.
19. Miller A, Chaung M, Teirstein AS, Siltzbach LE. Pulmonary function in stage I and II pulmonary sarcoidosis. *Ann NY Acad Sci* 1976; 278: 292–300.
20. Muller NL, Kullnig P, Miller RR. The CT findings of pulmonary sarcoidosis: analysis of 25 patients. *AJR Am J Roentgenol* 1989; 152: 1179–1182.
21. Mana J, Teirstein AS, Mendelson DS, et al. Excessive thoracic computed tomographic scanning in sarcoidosis. *Thorax* 1995; 50: 1264–1266.
22. Davies CW, Tasker AD, Padley SP, et al. Air trapping in sarcoidosis on computed tomography: correlation with lung function. *Clin Radiol* 2000; 55: 217–221.
23. Abehsera M, Valeyre D, Grenier P, et al. Sarcoidosis with pulmonary fibrosis: CT patterns and correlation with pulmonary function. *AJR Am J Roentgenol* 2000; 174: 1751–1757.
24. Paramothayan NS, Lasserson TJ, Jones PW. Corticosteroids for pulmonary sarcoidosis. *Cochrane Database Syst Rev* 2005; 18: CD001114.

25. Thomas KW, Hunninghake GW. Sarcoidosis. *JAMA* 2003; 289: 3300–3303.
26. Selroos O. Value of biochemical markers in serum for determination of disease activity in sarcoidosis. *Sarcoidosis* 1984; 1: 45–49.
27. Lieberman J, Nosal A, Schlessner A, Sastre-Foken A. Serum angiotensin-converting enzyme for diagnosis and therapeutic evaluation of sarcoidosis. *Am Rev Respir Dis* 1979; 120: 329–335.
28. NIH conference. Pulmonary sarcoidosis: a disease characterized and perpetuated by activated lung T-lymphocytes. *Ann Intern Med* 1981; 94: 73–94.
29. Lynch JP III. Computed tomographic scanning in sarcoidosis. *Semin Respir Crit Care Med* 2003; 24: 393–418.
30. Teirstein AS, Machac J, Almeida O, et al. Results of 188 whole-body fluorodeoxyglucose positron emission tomography scans in 137 patients with sarcoidosis. *Chest* 2007; 132: 1949–1953.
31. Sulica R, Teirstein AS, Kakarla S, et al. Distinctive clinical, radiographic, and functional characteristics of patients with sarcoidosis-related pulmonary hypertension. *Chest* 2005; 128: 1483–1489.
32. Foley RJ, Metersky ML. Successful treatment of sarcoidosis-associated pulmonary hypertension with bosentan. *Respiration* 2008; 75: 211–214.
33. Shorr AF, Davies DB, Nathan SD. Outcomes for patients with sarcoidosis awaiting lung transplantation. *Chest* 2002; 122: 233–238.

High-Resolution Computed Tomography Findings in Idiopathic Interstitial Pneumonias

Pamela A. Argiriadi, MD, and David S. Mendelson, MD

Mount Sinai School of Medicine, New York, NY

ABSTRACT

In June 2001, using an “international multidisciplinary consensus,” the American Thoracic Society and European Respiratory Society classified the idiopathic interstitial pneumonias (*Am J Respir Crit Care Med* 2002;165:277–304). The American Thoracic Society and European Respiratory Society coined the term *idiopathic interstitial pneumonias* because the diseases in question have both unknown etiologies and various combinations of inflammation and fibrosis within the lung parenchyma. The consensus included the development of a classification standard of the interstitial pneumonias and discussed the role of surgical lung biopsy and bronchoalveolar lavage evaluation. The classification of idiopathic interstitial pneumonias is based on histologic criteria, which correlate to imaging patterns on high-resolution computed tomography. The group of lung diseases under the umbrella of idiopathic interstitial pneumonias includes the following: usual interstitial pneumonia, nonspecific interstitial pneumonia, desquamative interstitial pneumonia, respiratory bronchiolitis interstitial lung disease, acute interstitial pneumonia, cryptogenic organizing pneumonia, and lymphoid interstitial pneumonia. *Mt Sinai J Med* 76:37–52, 2009. © 2009 Mount Sinai School of Medicine

Key Words: idiopathic, interstitial pneumonia, lung disease, high resolution CT.

The overall incidence of idiopathic interstitial pneumonias (IIPs) is not known (because of changes

in pathologic classification, older estimates are not valid), but they are thought to be rare. Raghu *et al.*¹ estimated the prevalence of idiopathic pulmonary fibrosis (IPF) in the overall US population to be 42.7 per 100,000 (using broad criteria) and 14.0 per 100,000 (using narrow criteria). However, thoracic manifestations of IIPs serve as prototypes for more frequent disorders, such as sarcoidosis and collagen vascular diseases (most commonly systemic sclerosis, rheumatoid arthritis, and systemic lupus erythematosus).²

The clinical presentation of IIPs is usually nonspecific. Exercise-limiting dyspnea with a nonproductive cough is the classic presentation. Gender predilection, age, and smoking history can suggest one IIP over another. Suspicion of a specific underlying etiology may be further refined on the basis of a clinical history, including medications and occupational and environmental exposure. On physical examination, basilar rales and clubbing can be seen in all types of IIPs. Cor pulmonale with elevated right atrial pressures may be seen in the late phase of any IIP. Pulmonary function tests (PFTs) usually show a restrictive lung pattern with decreased forced vital capacity, low resting partial pressure of oxygen in arterial blood (which falls even further with exercise), and decreased diffusion capacity of the lung for carbon monoxide. Xaubet *et al.*³ suggested that the forced vital capacity and diffusion capacity of the lung for carbon monoxide best reflect the diffuse extent of disease in IPF, providing possible information on disease progression.

Because clinical presentation and physical examination can be nonspecific in this patient population, high-resolution computed tomography (HRCT) has evolved into an easy, noninvasive tool that can evaluate the lung parenchyma on the millimeter scale, differentiate these entities from one another, and, more importantly, differentiate which cases will benefit from possible lung biopsy.

Address Correspondence to:

Pamela A. Argiriadi

Mount Sinai School of Medicine
New York, NY

Email: pamela.argiriadi@mountsinai.org

IMAGING OF IDIOPATHIC INTERSTITIAL PNEUMONIAS

Chest Radiographs

The most convenient and most inexpensive way to radiologically evaluate the thorax is the chest radiograph. However, many IIPs are radiographically occult. Epler *et al.*⁴ and Padley *et al.*⁵ discovered that at least 10% to 15% of patients with histologically proven infiltrative lung disease have normal chest radiographs. The classic radiographic findings in patients with IIPs include decreased lung volumes and increased interstitial markings. Radiographic findings of pulmonary arterial hypertension, including hilar enlargement, main pulmonary arterial enlargement, and right heart enlargement, are seen in longstanding disease. Of course, chest radiographs are invaluable in the patient with acute-on-chronic shortness of breath to assess for possible superimposed pneumonia, pneumothorax, congestive heart failure, or other acute processes.

High-Resolution Computed Tomography

The superiority of HRCT in disease detection and in distinguishing between different types of lung pathology has been well documented in the literature.⁶⁻⁸ HRCT remains the best imaging tool for evaluating the pulmonary interstitium and diagnosing interstitial lung disease.

At our institution, a computed tomography (CT) scan of the chest can be performed with either standard collimation (5 mm) or thin collimation (1- to 1.25-mm high-resolution protocol). The goal of thin collimation is to detect subtle abnormalities by an enhancement of spatial resolution. Diagnostic limitations at thinner collimation (>0.5 mm) occur because the signal-to-noise ratio decreases dramatically, and this limits detection.⁹ Diagnostic limitations also occur at thicker collimations (>2 mm) because of lung parenchymal overlap. Thin collimation images of the lungs are reconstructed with a high-spatial-frequency algorithm (otherwise known as a bone or lung algorithm) to enhance tissue interfaces.

Standard-dose parameters of HRCT include a peak kilovolt range of 110 to 130 and a milliamperage range of 200 to 300. With exposure times averaging 600 milliseconds, the milliamperage range is approximately 120 to 180. Short exposure times potentially limit artifact from respiratory motion. Because the lungs demonstrate an inherent high-contrast environment, low-dose CT parameters can be substituted for standard-dose parameters. In a study of 50 patients, Lee *et al.*¹⁰ found no significant difference between HRCT scans performed with 340 mA versus HRCT scans performed with 80 mA. Figure 1A,B presents 2 normal HRCT scans. One HRCT scan was performed with 270 mA, and the other HRCT scan was performed with 80 mA. The noise in the low-dose scan is apparent but does not limit diagnostic capability. However, low-dose scans are not for everybody. Caution should be

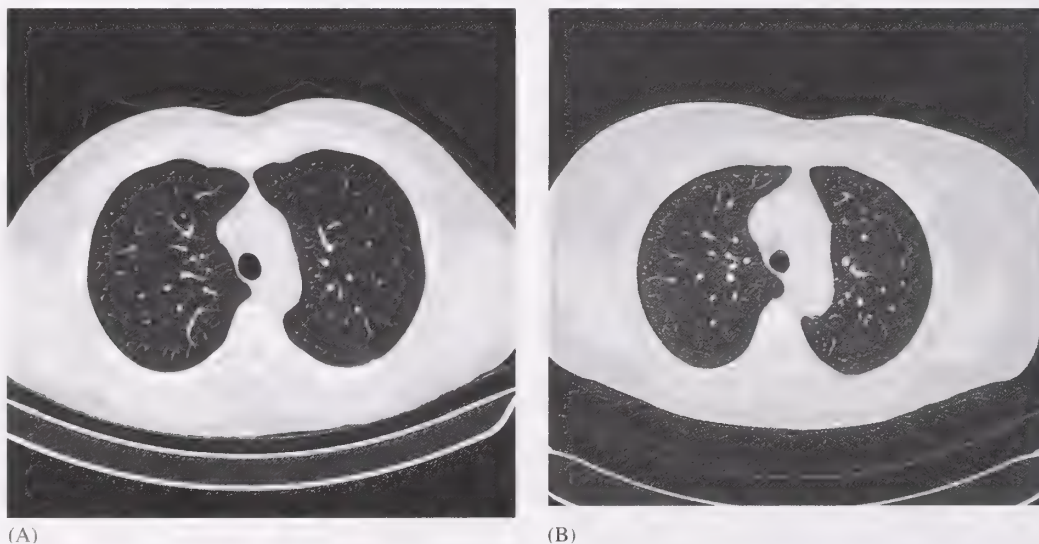


Fig 1. Visual difference between conventional and low-dose high-resolution chest computed tomography in a normal study. (A) The high-resolution computed tomography was performed with 270 mA. (B) The high-resolution computed tomography was performed with 100 mA. More noise is present in the high-resolution computed tomography scan on the right, but it does not preclude diagnosis.

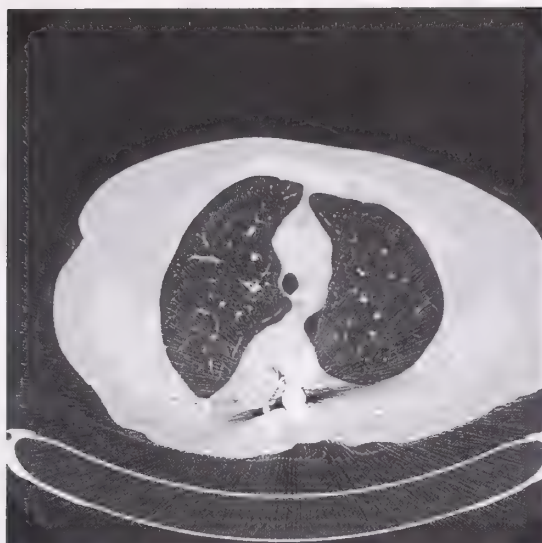


Fig 2. Beam-hardening artifact. This is a low-dose high-resolution computed tomography scan of a patient with thoracic spinal fixation hardware. There is extensive beam-hardening artifact, which obscures the posterior lung parenchyma.

used with morbidly obese patients, who have a substantial amount of internal scatter on any CT scan, and with patients that have metal devices (pacemakers, spinal fixation, humeral prostheses, etc.), which can obscure scan detail. Figure 2 presents a low-dose HRCT scan of a patient with a thoracic spinal fixation. The metallic fixation device obscures the lung parenchyma with extensive beam-hardening artifact in the thorax. This patient may have benefited from HRCT with a conventional dose.

Images may be obtained in a contiguous fashion, providing a full assessment of the lungs and mediastinum. However, another approach is to obtain thin sections at 1- to 2-cm intervals. This survey minimizes the radiation dose and can effectively demonstrate interstitial changes throughout all lung zones. One must appreciate that significant portions of the lung are excluded on this survey, and although this is effective for assessing the presence and distribution of interstitial disease, focal lesions may be missed.

Intravenous contrast can be added to better visualize the thoracic vascular structures, to assess for hilar lymphadenopathy, and to possibly assess for pulmonary nodule enhancement. A noncontrast examination can visualize the pulmonary interstitium and secondary pulmonary lobule just as accurately as a contrast-enhanced scan.

HRCT can be performed in both the inspiratory and expiratory phases and in both prone and supine positions. Inspiratory and expiratory scans are especially helpful in patients with possible air

trapping or bronchiolitis. Figure 3A,B presents an expiratory and inspiratory scan of the same patient at the level of the main pulmonary artery.

The lung volumes in the expiratory scan are decreased; there is also dependent ground glass in the posterior aspects of the right middle lobe and the bilateral lower lobes. These changes resolve on full inspiration. The expiratory image demonstrates how posterior atelectasis manifests as peripheral ground glass in patients, which can cause confusion in patients undergoing evaluation for IIPs. A full inspiration is paramount in evaluating this subset of patients. Otherwise, repeat imaging in the prone position can distinguish true ground-glass change from atelectasis, as atelectasis will appear in the most gravity-dependent portion of the lungs; hence, when the patient is in the prone position, such changes will vanish in the posterior lung fields and may be seen along the now dependent anterior portion of the lungs.

Finally, multidetector CT technology offers the ability to perform multiplanar reformation. In the realm of HRCT and in the evaluation of IIPs, coronal reconstruction is the most useful application of multiplanar reformation, as apical or basilar distribution can be identified more readily. Remy-Jardin *et al.*¹¹ demonstrated that coronal sections have diagnostic accuracy comparable to that of axial images on HRCT.

ANATOMIC CONSIDERATIONS AND BASIC TERMINOLOGY

Before delving into the different appearances of the IIPs, we will take a glance at standard lung imaging anatomy and basic terminology to facilitate further discussion (Table 1).

Müller¹² first described "the secondary pulmonary lobule as the smallest unit of lung structure margined by connective tissue septa," and it serves as the basic structure for lung imaging anatomy. Secondary pulmonary lobules are elegantly seen in HRCT, as demonstrated in Figure 4 for a patient with Niemann-Pick disease, in whom the interlobular septae are pronounced. The secondary pulmonary lobule is polyhedral in shape, measuring approximately 1.0 to 2.5 cm and containing a central bronchovascular bundle (usually about 1 mm thick), which includes a rich network of peribronchovascular interstitium. The walls of the polyhedron contain lymphatics and pulmonary venous branches. The secondary pulmonary lobule has on average 3 to 24 terminal bronchi¹² and contains thousands of alveoli.

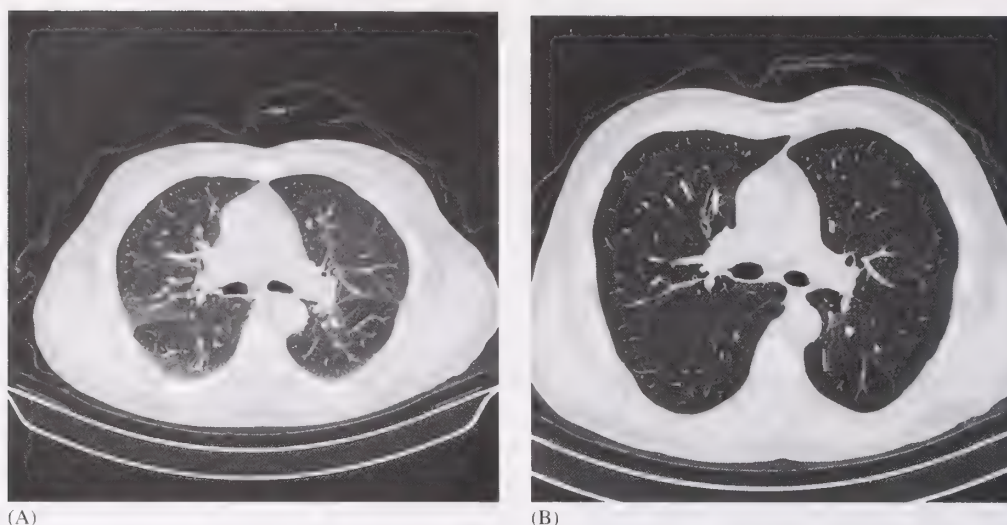


Fig 3. High-resolution computed tomography in the inspiratory and expiratory phases. (A) Expiratory-phase high-resolution computed tomography demonstrates ground glass in the dependent portions of the lungs, and (B) this resolves in inspiratory high-resolution computed tomography.

Table 1. Terminology.

Term	Definition
Consolidation (Figure 5A)	An increase in lung density that obscures underlying vessels. This indicates alveolar collapse or replacement of alveolar air with fluid or cells. Air bronchograms can be seen amidst consolidated lung.
Ground glass (Figure 5B)	A hazy increase in lung density that does not obscure underlying vessels and is nonspecific. This can represent partial collapse or partial filling of alveoli, interstitial inflammation, increased capillary blood volume, or a combination of these. This can be nodular, focal/patchy, or diffuse.
Mosaic perfusion	Geographic differences in lung attenuation associated with air trapping or oligemia. This is not to be confused with ground-glass opacity.
Honeycombing (Figure 6A)	Cystic airspaces with clearly defined fibrous walls. This results from or is associated with pulmonary fibrosis and loss of lung architecture.
Macrocysts	Honeycomb cysts greater than 1 cm.
Microcysts	Honeycomb cysts less than 1 cm.
Bronchiectasis (Figure 6B)	Bronchial dilatation. This is defined as a bronchial diameter greater than the diameter of the neighboring pulmonary arterial branch.
Traction bronchiectasis (Figure 6C)	Bronchial dilatation and bronchial irregularity in patients with pulmonary fibrosis. The traction of fibrotic lung tissue pulls on the bronchial walls.
Interlobular septal thickening (Figure 7A)	Abnormal thickening of interlobular septae, surrounding the secondary pulmonary lobule with fibrosis, edema, or cell infiltration.
Intralobular septal thickening or intralobular lines (Figure 7B)	Fine, meshlike reticulation within the secondary pulmonary lobule. This can be an early sign of fibrosis or lung infiltration.
Subpleural lines/reticulation (Figure 7C)	Thin opacity close to the pleural surface, usually paralleling the pleurae. This can represent atelectasis or fibrosis in the nondependent lung.
Parenchymal band (Figure 7D)	Linear opacities, several millimeters thick and several centimeters in length, that can occur anywhere within the lung but are usually in contact with the pleural surface. This is very commonly seen in scarring, atelectasis, or fibrosis.

NOTE: This table was created with data and terms from Chapter 12 of *High-Resolution CT of the Lung*.¹²

The IIPs, as a group, demonstrate various combinations of lung opacities, cystic air space disease, and reticulation. A brief vocabulary table (Table 1) illustrates terms that are helpful in elucidating the

findings on HRCT in IIP. These terms distinguish the IIPs from one another and from other interstitial lung diseases and help with correlations to pathology. Lung opacities are classified as either consolidations

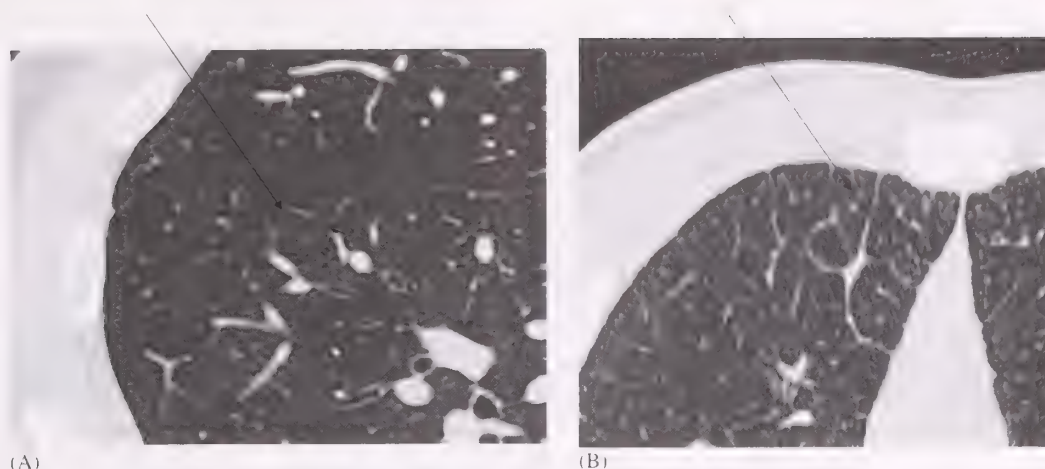


Fig 4. The secondary pulmonary lobule: (A) a normal secondary pulmonary lobule and (B) a secondary pulmonary lobule outlined with interlobular septal thickening in a patient with Niemann-Pick disease.

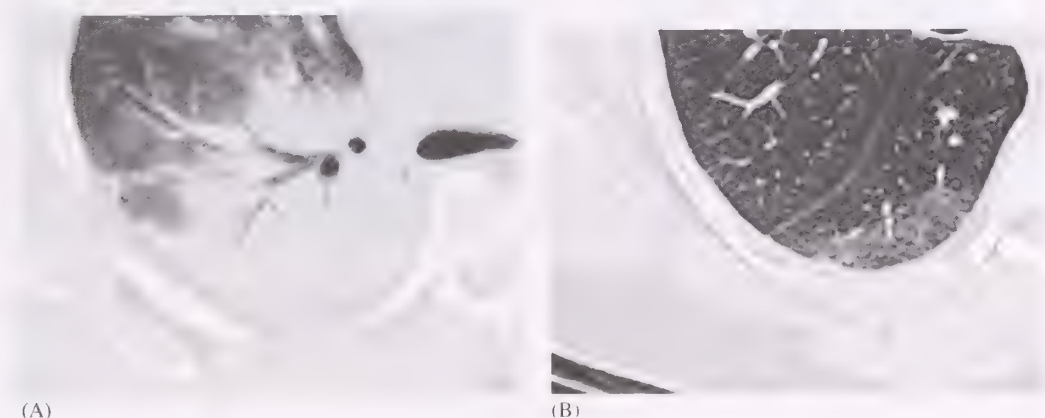


Fig 5. Lung opacities: (A) consolidation and (B) ground-glass opacity.

or ground glass (Figure 5). Cystic air space disease within the lung can include, but is not limited to, bronchiectasis, honeycombing, emphysematous change, and lung cysts/pneumatoceles (Figure 6). Reticular lung disease is divided into interlobular septal thickening, intralobular septal thickening, subpleural lines, and parenchymal bands (Figure 7).

IDIOPATHIC INTERSTITIAL PNEUMONIAS: CLINICAL, RADIOLOGICAL, AND PATHOLOGIC FEATURES

Idiopathic Pulmonary Fibrosis and Usual Interstitial Pneumonia

IPF is a distinctive type of chronic fibrosing pneumonia of unknown cause and is associated with a histologic pattern of usual interstitial pneumonia

(UIP).² The classic histological findings of IPF/UIP are fibroblastic foci with architectural distortion and honeycombing with intervening normal lung.^{2,9} The diagnosis of IPF is often a diagnosis of exclusion because the histologic appearance of UIP can occur on account of a variety of nonidiopathic causes, including collagen vascular diseases (eg, scleroderma, rheumatoid arthritis, and systemic lupus erythematosus), drug toxicity (eg, cytoxan and bleomycin), and environmental exposures (eg, asbestos and hypersensitivity pneumonitis).

Patients with IPF are usually older than 50 years and present with gradual, progressive dyspnea and nonproductive cough.¹³ There are more males than females diagnosed with IPF, but no definite gender predilection has been established.¹⁴ Patients with IPF commonly have a smoking history.¹⁴ On physical examination, signs of hypoxemia, including clubbing of the digits, occur in up to 50% of patients,¹³ and

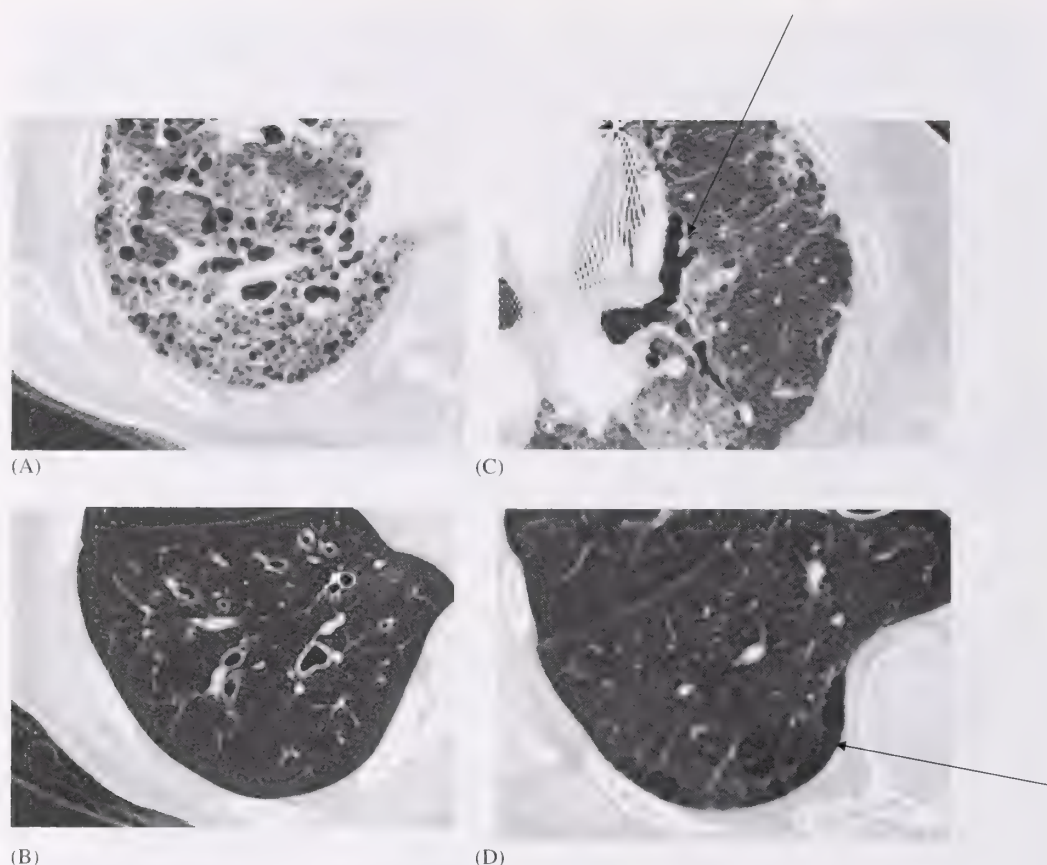


Fig 6. Cystic lung disease: (A) honeycombing, (B) bronchiectasis, (C) traction bronchiectasis, and (D) paraseptal emphysema.

chest auscultation reveals fine, Velcro-like rales at the lung bases.

Patients with IPF have a deteriorating course without a positive response to steroid treatment and with a median survival time of 2.5 to 3.5 years.² PFTs reflect its deteriorating course, as PFTs can initially be normal in patients with IPF but progressively demonstrate a restrictive pattern with decreasing partial pressure of oxygen in arterial blood.^{2,4} Patients without a smoking history appear to have a better prognosis than patients with a smoking history; however, patients with IPF that quit smoking appear to have a worse prognosis than patients that continue to smoke because of the healthy smoker effect.¹⁵

Because the prognosis of IPF/UIP is more severe in comparison with the prognosis of other IIPs, developing diagnostic criteria for IPF on HRCT is crucial.² The classic HRCT findings of IPF/UIP include intralobular septal thickening, subpleural reticular opacities, honeycombing, and lung fibrosis/architectural distortion, as demonstrated by traction bronchiectasis, in a peripheral-basilar distribution.^{2,9,16} Figure 8 demonstrates all of these

HRCT findings, which complement the histologic findings of UIP upon lung biopsy.

Honeycombing occurs in up to 80% of patients with IPF/UIP,⁹ usually in a macrocystic form.¹⁶ Ground-glass opacity is common in IPF/UIP, but it is not the dominant feature. Focal areas of ground glass are thought to represent lung fibrosis in evolution.⁹ However, if ground-glass opacity is the predominant finding and exceeds the extent of honeycombing within the lungs, other IIPs should be considered.² Often, the lung parenchyma demonstrates a variegated or heterogeneous appearance with fibrotic lung interspersed with normal lung.² The differential diagnosis of HRCT findings of IPF/UIP includes nonidiopathic interstitial pneumonias with a peripheral and basilar predominance, including asbestosis, chronic hypersensitivity pneumonitis, scleroderma, rheumatoid arthritis, and drug toxicity (eg, bleomycin and cytoxan). Figure 9 demonstrates peripheral honeycombing, subpleural lines, and intralobular septal thickening predominantly at the bases, which is classic for UIP and is confirmed with histology. Clinical correlation reveals that this is a 53-year-old woman with rheumatoid arthritis. Therefore, this case is an

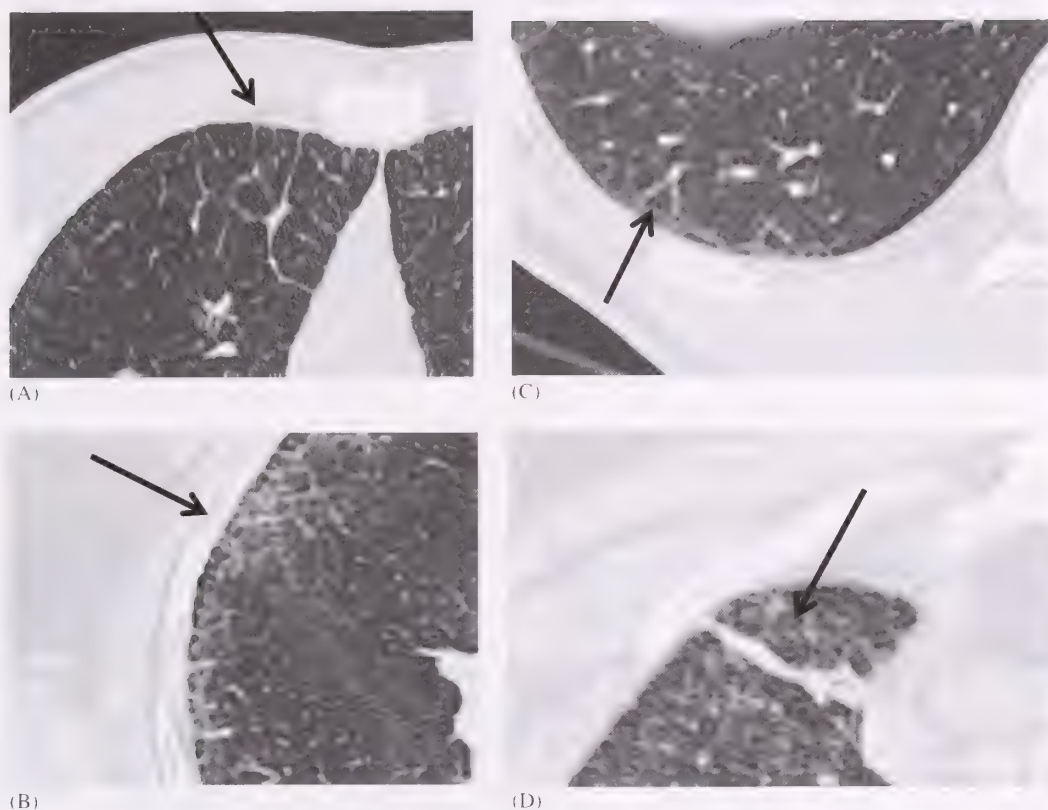


Fig 7. Reticular lung disease: (A) interlobular septal thickening, (B) intralobular septal thickening, (C) subpleural reticular opacities, and (D) parenchymal band.

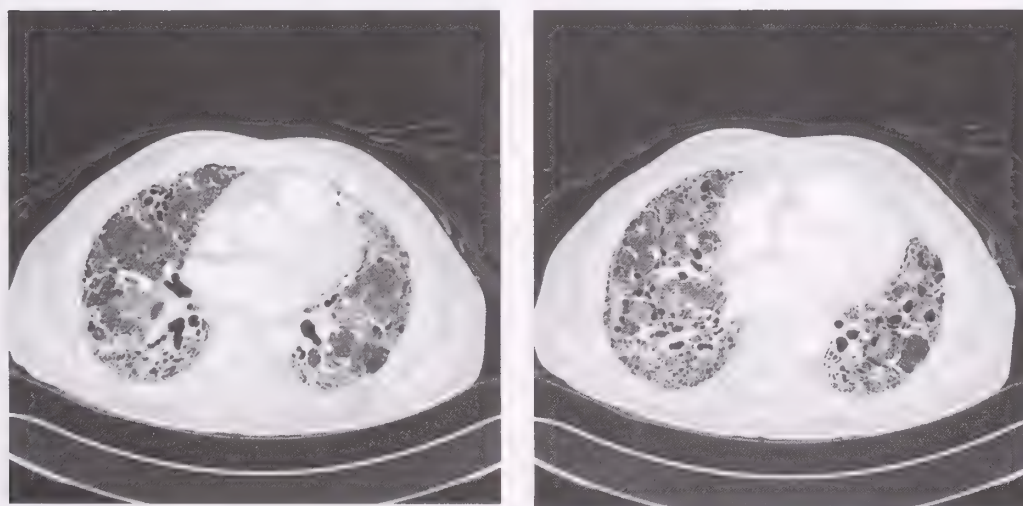


Fig 8. Idiopathic pulmonary fibrosis usual interstitial pneumonia: a patient with histologically proven usual interstitial pneumonia and clinically proven (or unproven) idiopathic pulmonary fibrosis. The overwhelming findings are honeycombing, which is worse at the lung periphery and at the lung bases, and traction bronchiectasis. The lung that appears normal is adjacent to the honeycomb lung, and this produces a heterogeneous appearance. Focal areas of ground glass are present on the left, but this is not the predominant finding.

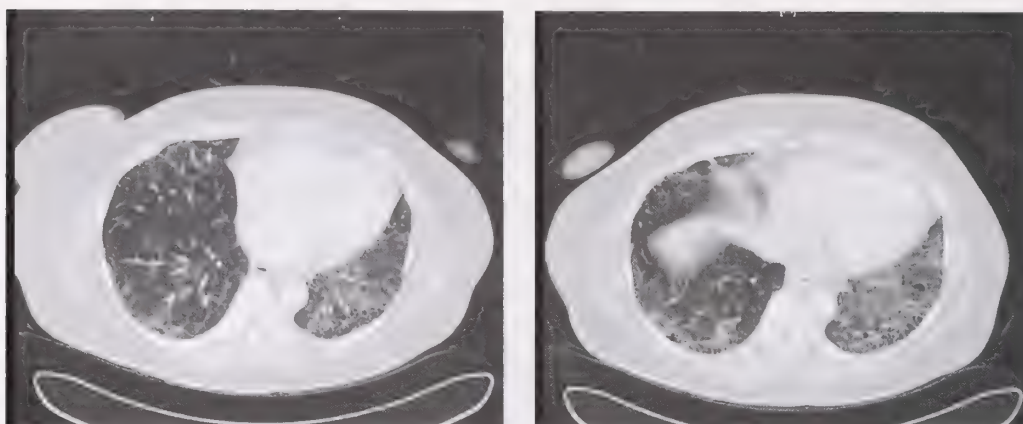


Fig 9. Usual interstitial pneumonia rheumatoid arthritis: histologically proven usual interstitial pneumonia in a patient with rheumatoid arthritis. The peripheral honeycombing is not as striking as that in the first case but nonetheless represents underlying architectural distortion.

example of UIP in rheumatoid arthritis and not IPF (Table 2).

Table 2. Usual Interstitial Pneumonia: Key Points.

Clinical findings
>50 years
Smoking history
Histologic findings
Honeycombing
Fibroblastic foci
Heterogeneous
High-resolution computed tomography findings
Macrocystic honeycombing (~ 80%)
Traction bronchiectasis
Peripheral distribution
Apical-basilar gradient
Clinical conditions associated with a histologic diagnosis of usual interstitial pneumonia
Idiopathic pulmonary fibrosis
Collagen vascular disease
Asbestosis
Hypersensitivity pneumonitis
Drug toxicity

Nonspecific Interstitial Pneumonia

The American Thoracic Society/European Respiratory Society consensus recognized that not all patients fulfill clinical and pathologic criteria for UIP, desquamate interstitial pneumonia (DIP), cryptogenic organizing pneumonia (COP), or acute interstitial pneumonia (AIP).² Katzenstein *et al.*¹⁷ wrote one of the first articles to describe an IIP "characterized by varying proportions of inflammation and fibrosis that appeared to be occurring over a single time span, i. e. temporally uniform" or homogeneous. These

patients were placed in a separate category called nonspecific interstitial pneumonia (NSIP) because they appeared to have better prognoses and different treatment options than their counterparts with UIP.¹⁷

Patients with NSIP are usually younger than patients with UIP (average age between 40 and 50 years), and unlike UIP, NSIP can occur in the pediatric population.² There is no gender predisposition or association with smoking.^{2,16} Clinical and physical findings of NSIP are similar to, but at times less severe than, the clinical findings of UIP.²

NSIP is histologically characterized by a homogeneous pattern of cellular interstitial inflammation with varying degrees of fibrosis.² Although a histologic sample with NSIP has a homogeneous appearance, NSIP is a very broad and inhomogeneous group. Pathologically, NSIP can be divided into cellular, fibrosing, or mixed subtypes.² Fibrosing subtypes of NSIP can be very difficult to distinguish from UIP.² Equally as important as temporal uniformity are pertinent negatives on the histologic specimen, including a lack of organizing pneumonia, fibroblastic foci, or diffuse alveolar damage. Many nonidiopathic clinical conditions are associated with an NSIP histologic pattern, such as collagen vascular disease, hypersensitivity pneumonitis, drug toxicity, and immunodeficiency.² With an NSIP histologic diagnosis, these etiologies should be excluded.

The clinical course of NSIP depends on the histologic subtype, with the fibrotic subtype exhibiting a poorer prognosis. According to Travis *et al.*,¹⁸ patients with a cellular pattern of NSIP had better 5- and 10-year survival than those with a fibrosing pattern of NSIP (100% versus 90% and 100% versus 35%). Overall, all patients with NSIP in that study had a better outcome than patients with UIP,

who had a 5-year survival of 43% and a 10-year survival of 15%.¹⁸

The most common HRCT feature of NSIP is ground-glass opacity, which occurs in 80% of cases and is usually subpleural and symmetric.^{2,9} Ground glass is the only visible abnormality in about 30% of cases, and this could reflect cases with the cellular type of NSIP.⁹ Honeycombing is rare, occurring in about 5% of cases.² When honeycombing does occur, it occurs typically in the microcystic form and predominantly in the fibrotic type of NSIP.¹⁶ More ground-glass opacity and honeycombing occur in the lower lobes than in the upper lobes, but the drastic apical-basilar gradient characteristically seen with UIP is not present.¹⁶ Intralobular septal thickening and subpleural lines occur in 50% of cases; consolidations

occur in approximately 25% of cases.^{2,9} Figures 10 to 12 present cases of histologically proven NSIP. The first case (Figure 10) demonstrates predominantly ground glass in the periphery of the lower lobes and right middle lobe; no other anomalies are present. Ground glass is still the predominant finding in the second case (Figure 11); however, subpleural reticular lines and microcystic honeycombing are present as well. The differential diagnosis of the first case, with the ground-glass predominance, includes cellular or mixed-type NSIP and DIP, whereas the differential diagnosis of the second case, with the honeycombing, includes fibrotic-type NSIP and UIP. The third case (Figure 12) demonstrates ground-glass opacity at the bases with intermixed microcystic honeycombing and intralobular septal thickening. An

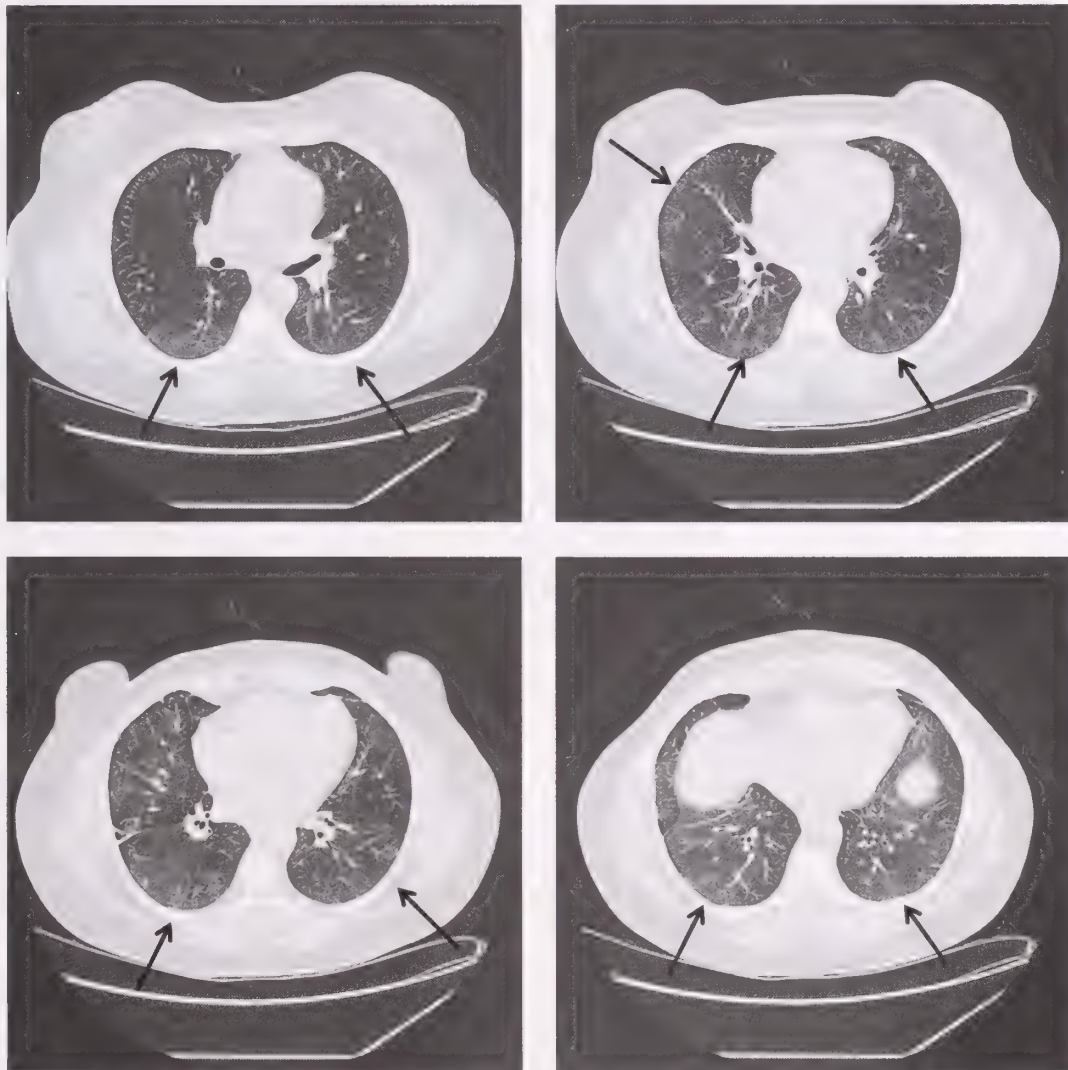


Fig 10. Nonspecific interstitial pneumonia. High-resolution computed tomography in a patient with nonspecific interstitial pneumonia demonstrates ground-glass opacity in the periphery of the lower lobes and right middle lobe. No honeycombing or reticular opacities are present. These findings could represent a cellular type of nonspecific interstitial pneumonia.

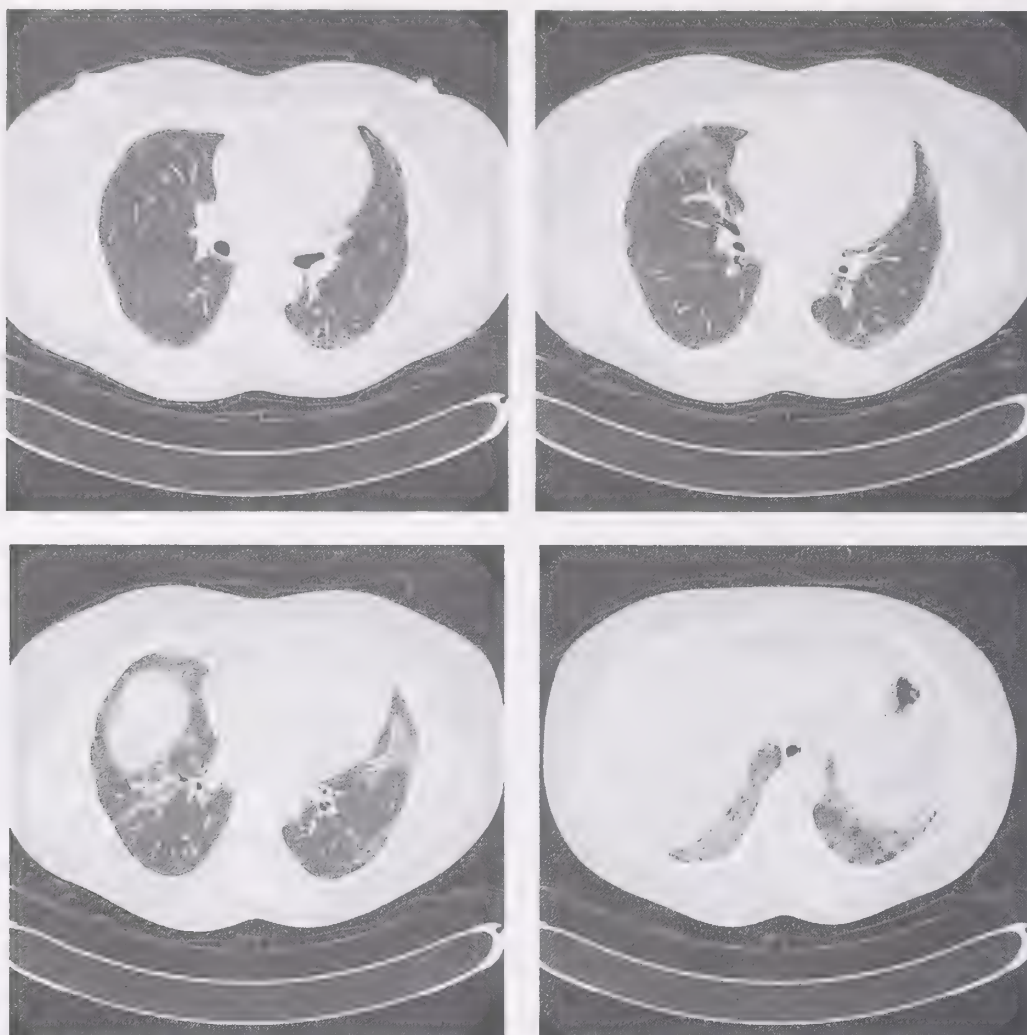


Fig 11. Nonspecific interstitial pneumonia. High-resolution computed tomography in a patient demonstrates ground-glass opacity in the anterior right middle lobe and lingula and in the posterior lower lobes. Microcystic honeycombing and subpleural lines are present at both costophrenic angles and are greater in the left lower lobe. These findings could represent a mixed or fibrotic type of nonspecific interstitial pneumonia.

evaluation of the mediastinal windows demonstrates a dilated, fluid-filled esophagus compatible with scleroderma, which is a clinical condition associated with an NSIP histologic diagnosis (Table 3).

Desquamative Interstitial Pneumonia and Respiratory Bronchiolitis Interstitial Lung Disease

The 2 smoking-related IIPs are respiratory bronchiolitis interstitial lung disease (RB-ILD) and DIP. There is significant overlap between the CT findings of respiratory bronchiolitis, RB-ILD, and DIP, since they represent different degrees of severity of airway and parenchymal reactions to cigarette smoke.¹⁸ Histologically, pigmented macrophages accumulate in the peribronchial interstitium and alveoli in RB-ILD and

throughout the lungs, without peribronchiolar preference, in DIP.²⁰

The average age of patients with both RB-ILD and DIP is 30 to 40 years, with men affected twice as much as women.¹⁶ Carrington *et al.*²⁰ observed that 36 of 40 patients (90%) with DIP had at least a 10-pack-year history of smoking. Clinically, these patients present with shortness of breath, cough, and restrictive lung symptoms. The prognosis of patients with RB-ILD and DIP is usually good with smoking cessation and corticosteroid therapy.¹⁶

Typical HRCT findings in patients with RB-ILD include bilateral ground-glass opacities in a patchy distribution with an upper lobe predominance and with intermixed centrilobular emphysema (given the history of smoking) and bronchial wall thickening.²¹

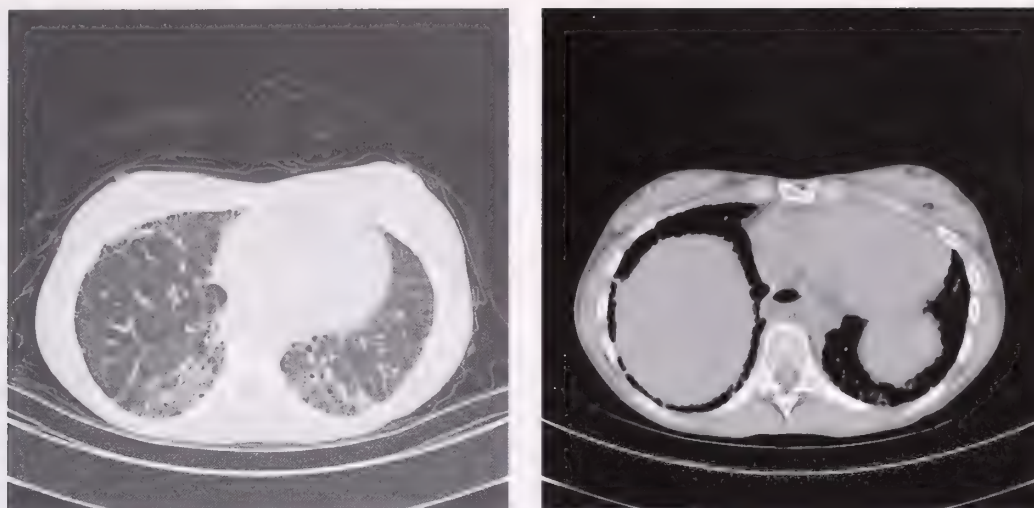


Fig 12. Nonspecific interstitial pneumonia in a patient with scleroderma. High-resolution computed tomography in a patient with histologically proven nonspecific interstitial pneumonia demonstrates ground-glass opacity, reticular opacities, and microcystic honeycombing. Mediastinal windows demonstrate a dilated and fluid-filled esophagus, which, upon correlation with clinical findings, confirmed the patient's diagnosis of scleroderma.

Table 3. *Nonspecific Interstitial Pneumonia—Key Points.*

Clinical findings
40–50 years
No smoking or gender predilection
Histologic findings
Pertinent negatives, excluding other idiopathic interstitial pneumonias, more important than pertinent positives
Three subtypes: cellular, fibrotic, and mixed
High-resolution computed tomography findings
Ground-glass predominance
Peripheral distribution
No obvious apical-basilar gradient
Microcystic honeycombing (~ 5%)
Reticular opacities (~ 50%)
Clinical conditions associated with nonspecific interstitial pneumonia histologic diagnosis
Collagen vascular disease (scleroderma ~ 40%)
Hypersensitivity pneumonitis
Drugs/toxins
Immunosuppression

Centrilobular nodules, likely indicating peribronchial inflammation, occur in 40% of patients.¹⁶ Subtle areas of fibrosis (honeycombing and intralobular lines) can occur in the periphery of the lung in up to 25% of patients.²¹

Typical HRCT findings in patients with DIP include bilateral, diffuse ground-glass opacity with a predisposition in the periphery (~ 60%) and the lower lobes (~ 100%); nonetheless, upper lobe disease is present in approximately 80% of cases

as well.²¹ The ground-glass opacity in both RB-ILD and DIP presumably represents the accumulation of alveolar macrophages within the lungs. Fibrosis, subtle areas of honeycombing, and intralobular lines occur more frequently in patients with DIP (up to 50%) versus patients with RB-ILD (up to 25%), and this indicates that DIP is the more severe lung disease on the continuum. Figure 13 presents HRCT images from a patient with lung biopsy demonstrating DIP and alveolar macrophage accumulation. This HRCT demonstrates the RB-ILD/DIP continuum. There is patchy peribronchial ground-glass opacity in the upper lobe (particularly the left upper lobe). At the lung bases, there is peripheral, denser ground-glass opacity with intermixed intralobular lines, indicating mild fibrosis (Table 4).

Cryptogenic Organizing Pneumonia

Organizing pneumonia is due to inflammation of the lung tissue with organization of exudates, rather than resorption, and with aggregation of granulation tissue polyps or buds within the distal air spaces on histologic evaluation.^{16,21} Organizing pneumonia is idiopathic, or cryptogenic (COP), in approximately 50% of cases.²² Noncryptogenic causes of organizing pneumonia include radiation therapy, drugs/toxins, and viral illnesses.² Conditions predisposing to COP include collagen vascular diseases, sarcoidosis, and immunosuppression.¹⁶

COP occurs equally in men and women, there is no smoking predilection, and the average age of patients with COP is 55 years.²³ Prognosis is

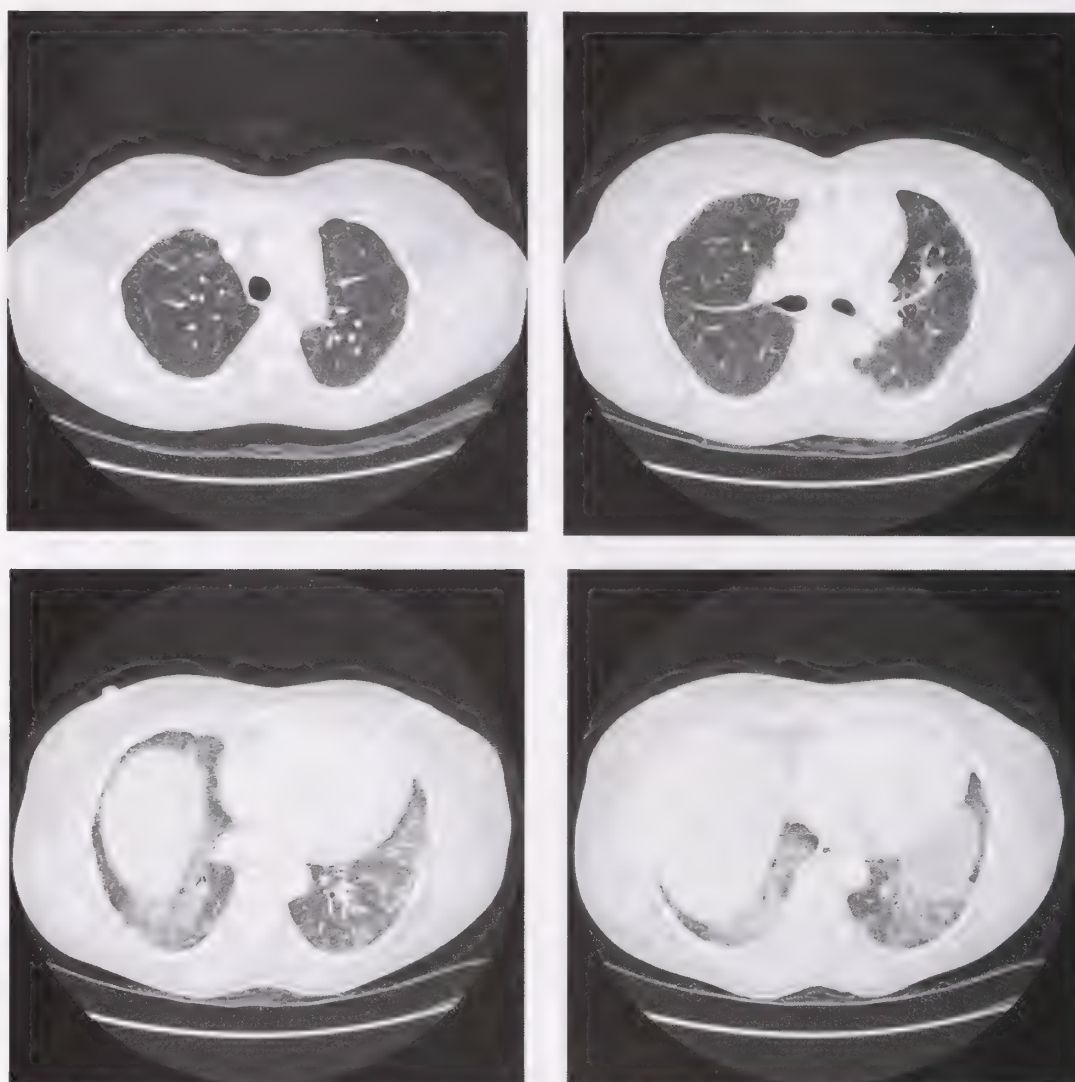


Fig 13. Desquamative interstitial pneumonia: high-resolution computed tomography from a patient with desquamative interstitial pneumonia on lung biopsy. The image on the top left demonstrates patchy peribronchial ground-glass opacity, whereas the next 3 images demonstrate peripheral ground-glass opacity with intermixed intralobular/subpleural lines; this is compatible with the disease spectrum, that is, respiratory bronchiolitis interstitial lung disease and desquamative interstitial pneumonia.

good; the majority of patients have a positive response to corticosteroids, but relapses are common within 3 months if steroid therapy is tapered or stopped.¹⁶

There are several different appearances of COP on HRCT. COP can present with patchy, peripheral ground-glass opacities in 60% of patients and with consolidations in 90% of patients.² These opacities and consolidations have lower lobe predominance and can be migratory, unilateral or bilateral, and symmetric or asymmetric.¹⁶ Air bronchograms almost always occur when consolidations are present, and intermixed small nodules (>10 mm) can occur in 50% of cases as well.² Figure 14 demonstrates asymmetric, patchy, peripheral opacities and consolidations with

intermixed small nodules in a patient with COP. In approximately 15% of patients, COP can present with irregularly shaped nodules with air bronchograms.² Figure 15 demonstrates 2 irregularly shaped nodules (one in the left upper lobe and one in the right lower lobe) also in a patient with COP. The differential diagnosis of HRCT findings of COP includes multilobar pneumonia in the more acute setting and bronchoalveolar carcinoma, alveolar sarcoidosis, and lymphoma in the more chronic setting. Chronic eosinophilic pneumonia should also be considered if the patient has peripheral waxing and waning opacities and/or consolidations (Table 5).

Table 4. *Desquamative Interstitial Pneumonia Respiratory Bronchiolitis Interstitial Lung Disease: Key Points.*

Clinical findings	
30–40 years	
>10-pack-year smoking history	2:1 men/women
Histologic findings	
Alveolar macrophages surrounding the bronchioles in respiratory bronchiolitis interstitial lung disease and diffusing throughout the lung in desquamative interstitial pneumonia	
High-resolution computed tomography findings in respiratory bronchiolitis interstitial lung disease	
Patchy ground-glass opacity and centrilobular nodules with upper lobe predominance	
High-resolution computed tomography findings in desquamative interstitial pneumonia	
Ground-glass opacity in the periphery of the lungs with lower lobe predominance	

Acute Interstitial Pneumonia

AIP is a fulminating disease, rapidly progressing to respiratory failure or death, and is histologically characterized as diffuse alveolar damage without an etiology.²⁴ Diffuse alveolar damage can be divided into an early exudative phase with interstitial edema and hyaline membrane accumulation, a proliferative phase with loose organizing fibrosis, and a fibrotic phase with ultimately honeycombing.² Katzenstein *et al.*²⁵ coined the term *acute interstitial pneumonia* in 1986, and it is synonymous with *Hamman-Rich lung*, which was first described in 1935.²⁶

AIP is histologically identical to acute respiratory distress syndrome (ARDS); however, AIP is idiopathic and occurs in patients with no previous history of lung disease. The clinical course of patients with AIP is also identical to that of patients with ARDS.²⁵ Patients have an acute onset (1–3 weeks) with cough and dyspnea and rapidly progress to respiratory failure with mechanical ventilation.²⁴ Steroid therapy can be effective in the earlier stage of AIP, but prognosis is poor with a 50% mortality rate.^{16,24} Men are equally as affected as women, there is no association with a smoking history, and the average age of a patient with AIP is 50 years.¹⁰ Clinical mimics of AIP include ARDS and diffuse alveolar damage from known causes, including sepsis, drug reactions, transfusion, and fulminant pulmonary manifestations of systemic diseases (eg, collagen vascular disease and vasculitides).²⁴ An acute exacerbation of UIP/IPF can also mimic AIP.²⁴

HRCT findings of AIP include extensive bilateral ground-glass opacities in a diffuse or patchy distribution with focal sparing; consolidations are seen less frequently.^{2,21} There is a lower lobe and posterior predominance.¹⁶ The presence of traction bronchiectasis and honeycombing determines the phase of the disease. If traction bronchiectasis is present, the ground-glass attenuation represents alveolar fibrosis and thus a late proliferative or fibrotic stage of AIP. If traction bronchiectasis is not present, the ground-glass attenuation represents alveolar edema and hyaline membrane accumulation and thus an exudative or early proliferative phase. Figure 16 presents an HRCT scan of a patient with AIP. There is fibrotic change in the peripheral and posterior aspect

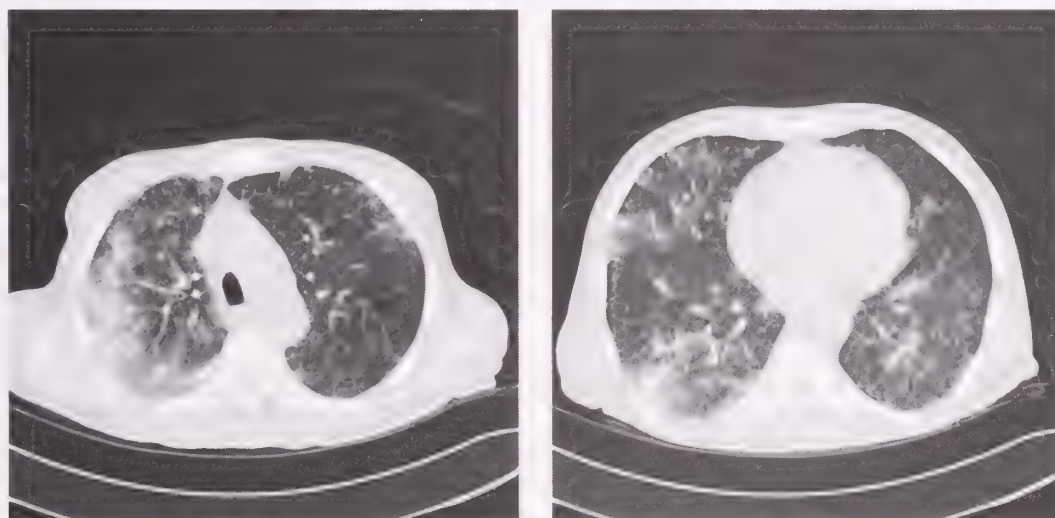


Fig 14. Cryptogenic organizing pneumonia. High-resolution computed tomography demonstrates asymmetric, patchy, peripheral opacities and consolidations with intermixed small nodules in a patient with cryptogenic organizing pneumonia.

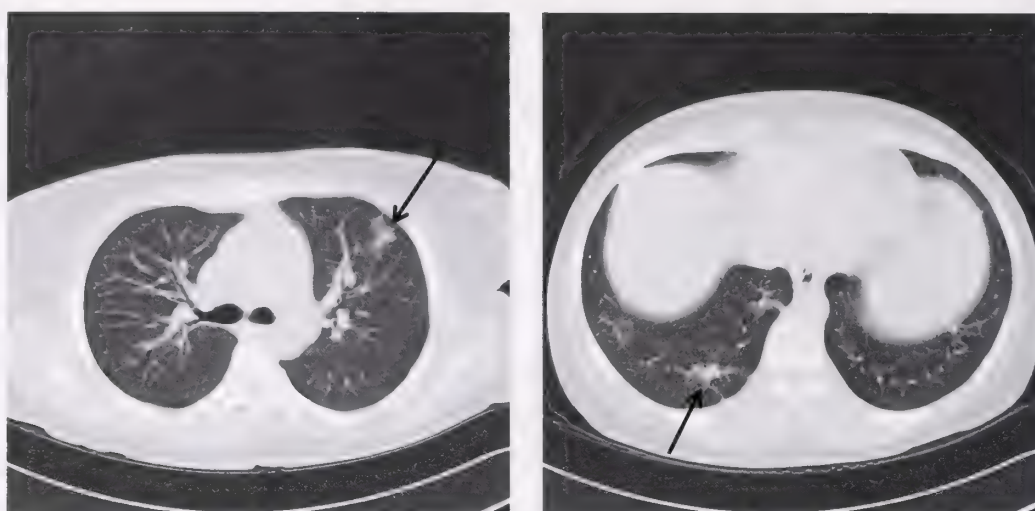


Fig 15. Cryptogenic organizing pneumonia. High-resolution computed tomography demonstrates 2 irregularly shaped nodules (one in the left upper lobe and one in the right lower lobe) in a patient with cryptogenic organizing pneumonia.

Table 5. *Cryptogenic Organizing Pneumonia: Key Points.*

Clinical findings
Average age = 55 years
Men = women
Histologic findings
Granulation polyps in the distal airspaces
High-resolution computed tomography findings
Patchy peripheral waxing/waning consolidations or opacities
Nodules or consolidations with air bronchograms
Clinical conditions associated with cryptogenic organizing pneumonia
Collagen vascular disease
Immunosuppression
Drugs/toxins
Radiation
Viral illness

of the lung, but with superimposed ground glass. The differential diagnosis of AIP also depends on the presence of honeycombing or traction bronchiectasis. If there is no evidence of fibrosis, the differential diagnosis includes multilobar pneumonia, pulmonary hemorrhage, pulmonary edema, and ARDS. If HRCT demonstrates fibrotic change, the differential diagnosis includes ARDS, acute exacerbation of UIP, and fibrotic-type NSIP (Table 6).

Lymphoid Interstitial Pneumonia

Lymphoid interstitial pneumonia (LIP) is a very rare IIP and is more common as an association with other systemic disorders (eg, Sjögren syndrome, other autoimmune disorders, and immunosuppression).¹⁶ Histologically, LIP is characterized by interstitial

infiltration with lymphocytes, plasma cells, and histiocytes. For this reason, some physicians consider LIP more of a lymphoproliferative disorder rather than a true IIP.² LIP is more common in women, likely because of the correlation with autoimmune disorders.² LIP is also present in the pediatric population as it represents 22% to 75% of pediatric patients with both human immunodeficiency virus and pulmonary disease.²⁷ HRCT findings in patients with LIP include diffuse ground-glass opacity, sometimes with a lower lobe predominance with intermixed central or perivascular cysts.¹⁶

CONCLUSION

The American Thoracic Society/European Respiratory Society consensus in 2001 included clinical, radiologic, and pathologic criteria for the 7 different types of IIPs. This consensus provided excellent imaging-pathology correlation for not only all the IIPs but also certain subtypes of NSIP and AIP. HRCT is the definitive imaging tool for deciphering the presence and individual findings of interstitial lung disease. However, whether HRCT can provide adequate diagnostic information to establish a final diagnosis depends tremendously on a detailed clinical history because IIPs are associated with many clinical syndromes. Although in general a differential diagnosis is often based on a combination of clinical, laboratory, radiological, and pathological findings, nowhere is this truer than in the diagnosis of IIP. Although pathologic findings are often considered the gold standard in most diseases, in the realm

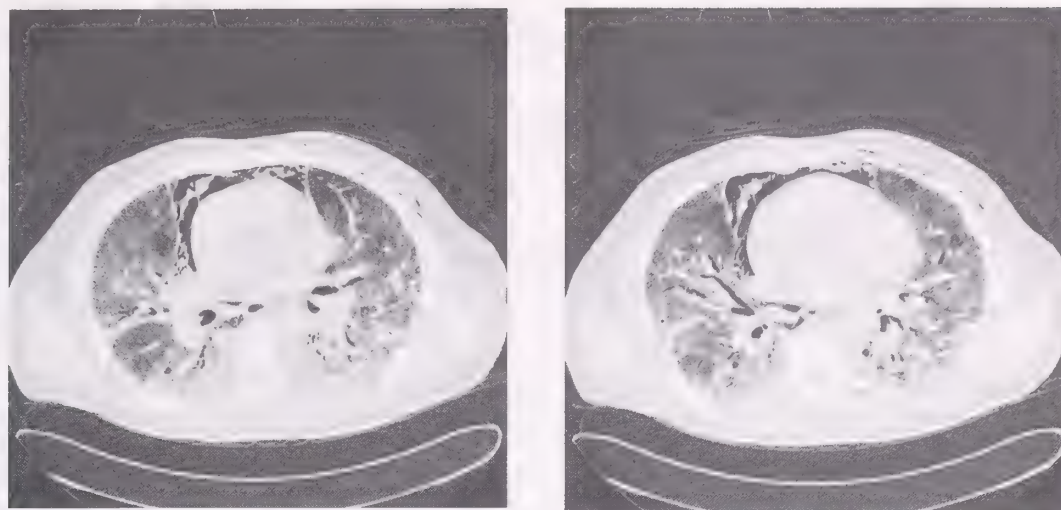


Fig 16. Acute interstitial pneumonia (or Hamman-Rich lung). A high-resolution computed tomography scan of a patient with acute interstitial pneumonia demonstrates fibrotic changes (traction bronchiectasis and honeycombing) in the peripheral and posterior aspect of the lung, but with superimposed ground glass. Incidentally, there is also pneumomediastinum as well as subcutaneous emphysema, likely secondary to ventilation. This case could represent acute interstitial pneumonia due to exacerbation of usual interstitial pneumonia or idiopathic acute interstitial pneumonia.

Table 6. *Acute Interstitial Pneumonia: Key Points.*

Clinical findings	
Average age = 50 years	
Men = women	
Histologic findings	
Diffuse alveolar damage, which is divided into early exudative and proliferative fibrotic phases	
High-resolution computed tomography findings	
Extensive posterior and basilar ground-glass opacities	
Presence of fibrosis with ground-glass opacity, which can help determine the phase of acute interstitial pneumonia	
Clinical conditions associated with diffuse alveolar damage	
Acute respiratory distress syndrome (sepsis, drug reaction, and transfusion)	
Collagen vascular disease	
Vasculitides	

of IIP, pathologic findings do not stand alone in reaching a diagnosis; rather, it is the combination of pathology and the preceding list that provides for the diagnosis in the current era. Particularly important clinical information includes a history of immunosuppression, drug toxicity, inhalational exposures, collagen vascular diseases, and other pertinent systemic syndromes.

Furthermore, HRCT should decrease the need for lung biopsy in certain IIPs. According to Raghu *et al.*,²⁸ the diagnosis of UIP can be made in patients without lung biopsy and with appropriate clinical and HRCT features. UIP and AIP should be

distinguished radiologically from other IIPs because of their poorer prognoses. Ground-glass opacity can be a very confounding variable on HRCT. Although it is a very sensitive finding showing that a disease process is present, it can represent a number of underlying causes ranging from partial alveolar filling with edema, tumor cells, hemorrhage, or pneumonia to fibrosis in evolution in UIP and in fibrotic-types of NSIP and AIP to granulation tissue in COP, cellular infiltration in NSIP, and macrophage accumulation in DIP. The key to making any diagnostic headway is to assess the presence, if any, of fibrosis (macrocytic honeycombing, microcytic honeycombing, and traction bronchiectasis). If the amount of fibrosis exceeds the amount of ground-glass opacity and the ground glass occupies less than 30% of the lung, it is safe to make a diagnosis of UIP. If ground-glass opacity exceeds the degree of honeycombing, lung biopsy should be performed to achieve a final diagnosis.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Raghu G, Weycker D, Edelsberg J, et al. Incidence and prevalence of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2006; 174: 810–816.

2. American Thoracic Society, European Respiratory Society. International multidisciplinary consensus classification of the idiopathic interstitial pneumonias [published correction appears in *Am J Respir Crit Care Med* 2002;166:426]. *Am J Respir Crit Care Med* 2002; 165: 277–304.
3. Xaubet A, Agustí C, Luburich P, et al. Pulmonary function tests and CT scan in the management of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1998; 158: 431–436.
4. Epler GR, McCloud TC, Gaensler EA, et al. Normal chest roentgenograms in chronic diffuse infiltrative lung disease. *N Engl J Med* 1978; 298: 934–939.
5. Padley SP, Hansell DM, Flower CD, Jennings P. Comparative accuracy of high resolution computed tomography and chest radiography in the diagnosis of chronic diffuse infiltrative lung disease. *Clin Radiol* 1990; 44: 227–231.
6. Müller NL. Computed tomography and magnetic resonance imaging: past, present, and future. *Eur Respir J* 2002; 35(suppl): 3–12.
7. Kazerooni EA. High-resolution CT of the lungs. *AJR Am J Roentgenol* 2001; 177: 501–519.
8. Mathieson JR, Mayo JR, Staples CA, Müller NL. Chronic diffuse infiltrative lung disease: comparison of diagnostic accuracy of CT and chest radiography. *Radiology* 1989; 171: 111–116.
9. Zompatori M, Sverzellati N, Poletti V, et al. High-resolution CT in diagnosis of diffuse infiltrative lung disease. *Semin Ultrasound CT MRI* 2005; 26: 332–347.
10. Lee KS, Primack SL, Staples CA, et al. Chronic infiltrative lung disease: comparison of diagnostic accuracies of radiography and low- and conventional-dose thin section CT. *Radiology* 1994; 191: 669–673.
11. Remy-Jardin M, Campistron P, Amara A, et al. Usefulness of coronal reformations in the diagnostic evaluation of infiltrative lung disease. *J Comput Assist Tomogr* 2003; 27: 266–273.
12. Webb WR, Müller NL, Naidich DP. *High-Resolution CT of the Lung*. 3rd ed. Philadelphia, PA: Lippincott, Williams & Wilkins; 2001.
13. Turner-Warwick M, Burrows B, Johnson A. Cryptogenic fibrosing alveolitis: clinical features and their influence on survival. *Thorax* 1980; 35: 171–180.
14. Coultas DB, Zumwalt RE, Black WC, Sobonya RE. The epidemiology of interstitial lung diseases. *Am J Respir Crit Care Med* 1994; 150: 967–972.
15. Antoniou KM, Hansell DM, Rubens MB, et al. Idiopathic pulmonary fibrosis: outcome in relation to smoking status. *Am J Respir Crit Care Med* 2008; 177: 190–194.
16. Mueller-Mang C, Grosse C, Schmid K, et al. What every radiologist should know about idiopathic interstitial pneumonias. *Radiographics* 2007; 27: 595–615.
17. Katzenstein AL, Fiorelli RF. Nonspecific interstitial pneumonia/fibrosis. Histologic features and clinical significance. *Am J Surg Pathol* 1994; 18: 136–147.
18. Travis WD, Matsui K, Moss J, Ferrans VJ. Idiopathic nonspecific interstitial pneumonia: prognostic significance of cellular and fibrosing patterns. Survival comparison with usual interstitial pneumonia and desquamative interstitial pneumonia. *Am J Surg Pathol* 2000; 24: 19–33.
19. Heyneman LE, Ward S, Lynch DA, et al. Respiratory bronchiolitis, respiratory bronchiolitis-associated interstitial lung disease, and desquamative interstitial pneumonia: different entities or part of the spectrum of the same disease process? *AJR Am J Roentgenol* 1999; 173: 1617–1622.
20. Carrington CB, Gaensler EA, Coutu RE, et al. Natural history and treated course of usual and desquamative interstitial pneumonia. *N Engl J Med* 1978; 298: 801–809.
21. Schaefer-Prokop C, Prokop M, Fleischmann D, Herold C. High resolution CT of diffuse interstitial lung disease: key findings in common disorders. *Eur Radiol* 2001; 11: 373–392.
22. Epler GR, Colby TV, McCloud TC, et al. Bronchiolitis obliterans organizing pneumonia. *N Engl J Med* 1985; 312: 152–158.
23. Cordier JF. Organizing pneumonia. *Thorax* 2000; 55: 318–328.
24. Bouros D, Nicholson AC, Polychronopoulos V, du Bois RM. Acute interstitial pneumonia. *Eur Respir J* 2000; 15: 412–418.
25. Katzenstein AL, Myers JL, Mazur MT. Acute interstitial pneumonia: a clinicopathologic ultrastructural and cell kinetic study. *Am J Surg Pathol* 1986; 10: 256–267.
26. Hamman L, Rich A. Fulminating diffuse interstitial fibrosis of the lungs. *Trans Am Clin Climatol Assoc* 1935; 51: 154–163.
27. Becciolini V, Gudinchet F, Cheseaux JJ, Schnyder P. Lymphocytic interstitial pneumonia in children with AIDS: high-resolution CT findings. *Eur Radiol* 2001; 11: 1015–1020.
28. Raghu G, Mageto YN, Lockhart D, et al. The accuracy of the clinical diagnosis of new-onset idiopathic pulmonary fibrosis and other interstitial lung disease: a prospective study. *Chest* 1999; 116: 1168–1174.
29. Webb RW. Thin-section CT of the secondary pulmonary lobule: anatomy and the image—the 2004 Fleischner lecture. *Radiology* 2006; 239: 332–338.
30. Ryu JH, Colby TV, Hartman TE. Idiopathic pulmonary fibrosis: current concepts. *Mayo Clin Proc* 1998; 73: 1085–1101.
31. Hidalgo A, Franquet T, Giménez A, et al. Smoking-related interstitial lung disease: radiologic-pathologic correlation. *Chest* 2006; 16: 2463–2470.

Comorbidities of Advanced Lung Disease

Joel Anthony Nations, MD,¹ and Steven D. Nathan, MD²

¹Pulmonary and Critical Care Medicine, National Naval Medicine Center, Bethesda, MD

²Inova Fairfax Hospital, Falls Church, VA

ABSTRACT

Lung disease in advanced stages is often accompanied by multiple comorbidities. Often, these processes lead to significant morbidity and occasionally mortality. Symptoms of pulmonary comorbidities, including pulmonary embolus, pulmonary hypertension, sleep-disordered breathing, and lung malignancy, may be intertwined with symptoms of advancing lung disease. However, the occurrence of extrapulmonary processes, including depression, diabetes mellitus, osteoporosis, and gastroesophageal reflux disease, may be subtle. Clinicians caring for advanced lung disease should be aware of the many comorbid processes that may accompany advancing lung disease. *Mt Sinai J Med* 76:53–62, 2009. © 2009 Mount Sinai School of Medicine

Key Words: Advanced lung disease, comorbidity, IPF, COPD.

Advanced lung disease is a term used to denote severe lung disease for which symptoms limit daily activity and therapy is limited. Patients are often plagued by additional interrelated illnesses that increase the mortality and morbidity of the underlying pulmonary process. Therefore, medical care of these disorders should entail a systemic approach. Understanding the spectrum and nature of these comorbid illnesses aids in developing comprehensive treatment plans and is therefore important to physicians who treat

patients with the various forms of advanced lung disease.

Common lung diseases that may progress to advanced lung disease include idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), sarcoidosis, cystic fibrosis (CF), and pulmonary arterial hypertension (PAH). However, any progressive lung disease can become advanced. Each of these diseases can have diverse presentations and clinical variances. The spectrum of comorbidities can be subdivided into pulmonary and extrapulmonary comorbidities, as delineated in Table 1. Distinguishing what constitutes a comorbidity from the clinical manifestation of a lung disease can be inexact because there may be overlap between a pulmonary comorbidity and the underlying parenchymal process. Parenchymal and airway destruction may increase susceptibility to pulmonary infections, including bacterial pneumonia and pulmonary aspergilloma. Additionally, pulmonary vascular processes such as pulmonary embolus and pulmonary hypertension (PH) may accompany destructive parenchymal disease. Therefore, although symptoms of dyspnea and cough commonly accompany advanced lung disease, they may also herald a pulmonary comorbidity. For example, cough might be a symptom of a superimposed comorbid lung cancer. In the case of dyspnea, this might occur in conjunction with the aforementioned pulmonary embolus or PH. Although these processes could be considered to be within the continuum of severe lung disease, there are other processes that may accompany advanced lung disease and are more distinct comorbid processes. Extrapulmonary comorbidities are important factors for patients with advanced lung disease as they may add morbidity and often herald increased mortality. Such common comorbidities include coronary artery disease (CAD), depression, gastroesophageal reflux disease (GERD), osteoporosis, and deconditioning.

Address Correspondence to:

Joel Anthony Nations

Pulmonary and Critical Care
Medicine

National Naval Medicine Center
Bethesda, MD

Email: joel.nations@med.navy.mil

Table 1. *Pulmonary and Extrapulmonary Comorbidities*

Pulmonary Comorbidities
Pulmonary parenchyma
• Bronchogenic carcinoma
• Infections
Pulmonary vasculature
• Pulmonary embolus
• Pulmonary hypertension
Obstructive sleep apnea
Extrapulmonary Comorbidities
Cardiac
• Coronary artery disease
• Heart failure
Endocrine
• Hypogonadism
• Diabetes mellitus
Gastroesophageal reflux
Psychiatric
• Depression
• Anxiety
Musculoskeletal
• Deconditioning
• Osteoporosis
Medication side effects

PULMONARY PARENCHYMA

Lung Cancer

Bronchogenic carcinoma is seen commonly in patients with fibrotic and chronic obstructive lung disease. In a population-based cohort study of 890 patients with IPF and 5884 control patients from the UK General Practice Research Database, the incidence of lung cancer was markedly increased among patients with IPF¹ with a relative risk ratio of 7.31. Adjusting for previous smoking history did not significantly alter this ratio. Bronchogenic carcinoma was the cause of death in 10.4% of patients with IPF as reported by Panos.² Whether the predisposing factors leading to bronchogenic carcinoma are squamous metaplasia, atypical epithelial cells, cigarette smoking, or other exposures is unknown. In a retrospective review of 24 patients with IPF and lung cancer, it was shown that IPF-associated lung cancer commonly arose in the lower lobes (42%) and was mainly composed of squamous cell carcinoma (67%).³ The mean survival in patients with IPF was 2.3 years after the diagnosis was made and 1.6 years after lung cancer was diagnosed. There was a 7:1 male-to-female ratio in the IPF and carcinoma group versus a 1:1 ratio in the IPF-without-carcinoma group, and this suggests that men with IPF have a higher risk for lung cancer. Interestingly, 42% of the cases of lung cancer in IPF patients were incidentally found during routine radiographic evaluation of lung disease. Patients in

this review had a mean age of 72.3 years and were predominantly current or past smokers (91%).³

Historically, sarcoidosis has also been implicated as a risk factor for lung cancer.⁴ However, in a study comparing lung cancer rates in patients with sarcoidosis and a population-based control from the Danish Cancer Registry, there was not a statistically significant increase in the rate of lung cancer in these patients.⁵ Confounding factors in establishing a relationship between sarcoidosis and lung cancer include the unknown prevalence of sarcoidosis and the presence of sarcoid-like granulomas that can be seen accompanying lung cancer.^{4,5} Controversy will likely continue concerning the true relationship of sarcoidosis and lung cancer.

Cigarette smoking is a major risk factor for both bronchogenic carcinoma and COPD. Whether there is increased risk of lung cancer development in patients with COPD regardless of smoking has been debated in the past. However, clear evidence now suggests that COPD even in the absence of smoking carries an increased risk of lung cancer.^{6,7}

Patients with advanced lung disease may have a limited range of treatment options for bronchogenic carcinoma because thoracic radiation and resection may not be appropriate. A multidisciplinary approach including clinicians adept at advanced lung disease is paramount. This should also frequently include palliative or hospice care experts, especially toward the end of the patient's disease course.

Infections

Infections often complicate advanced lung disease. Interestingly, in a retrospective review of patients with IPF admitted to the intensive care unit, pneumonia (31%) was second only to IPF exacerbation (47%) as the leading cause of respiratory failure.⁸ Whether this was due to underlying parenchymal damage or treatment with immunosuppressive agents including corticosteroids is unclear. Distinguishing infections from exacerbations of IPF is important. Additionally, because of previous hospitalizations and immunosuppression, these patients may be at risk for less common pathogens including fungi and *Pseudomonas* species. Aspergilloma (Figure 1) has been described as an important risk factor for increased mortality in patients with sarcoidosis.⁹

PULMONARY VASCULATURE

Pulmonary embolism (Figure 2) occurs with increasing rates in those patients with risk factors such as malignancy, immobility, heart failure, surgery, and

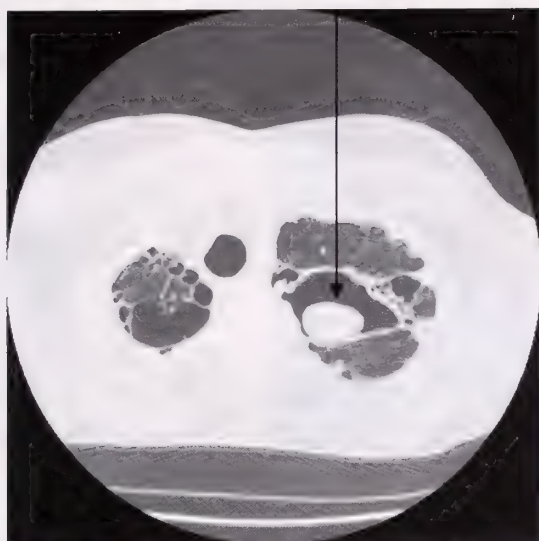


Fig 1. A large aspergilloma in the fibrotic left lung of a patient with advanced lung disease.

thrombophilia. Patients with advanced lung disease may possess multiple risks. For example, as with most forms of advanced lung disease, patients with IPF often have decreased mobility due to severe exertional dyspnea.¹⁰ Some evidence suggests that IPF may also be associated with thrombophilia, including elevated factor VIII levels and antiphospholipid antibodies.¹¹ In 1 review, death was attributed to pulmonary embolus in 3.4% of patients with IPF.² Furthermore, in a retrospective review of IPF lung transplant recipients, pulmonary embolism was found to occur in 27% of cases.¹² Just as for pretransplant patients, many possible explanations for this increased incidence exist, including immobility, corticosteroid use, malignancy, and thrombophilia. Empiric treatment with anticoagulation for IPF remains controversial. Kubo *et al.*¹³ prospectively randomized patients with IPF to corticosteroids or combination therapy with corticosteroids and oral anticoagulation. Mortality from acute exacerbations of IPF was lower in the arm combining corticosteroids and anticoagulation treatment (18% versus 71% for corticosteroids alone). Additionally, the 3-year survival of 63% with corticosteroids plus anticoagulation was significantly better than the 3-year survival in the group with corticosteroids alone (35%). However, as no literature supports the benefit of anticoagulation alone, it is unclear where anticoagulation fits in the treatment armamentarium for IPF.

Pulmonary microvascular thrombus is a recognized pathologic feature of idiopathic pulmonary arterial hypertension (IPAH). This likely is the result of *in situ* thrombosis.¹⁴ Frank *et al.*¹⁵ retrospectively reviewed patients with primary PH

[currently classified as World Health Organization (WHO) group I: IPAH] and anorexigen-induced PH with respect to anticoagulant treatment. A significant mortality benefit was associated with anticoagulation therapy in comparison with the control group (median survival of 7.2 years versus 4.9 years). This was more significant for those with anorexigen-associated PH. Two other studies have also demonstrated that anticoagulation has a potential mortality benefit for patients with primary PH.^{14,16} Current guidelines therefore advocate anticoagulation for patients with IPAH.¹⁷ Empiric anticoagulant treatment for other forms of PH is of uncertain benefit and thus should be individualized for each patient.

Although PAH can lead to advanced lung disease, increased pulmonary pressure can also accompany other forms of advanced parenchymal lung disease.¹⁸ Recently, WHO categorized PH as shown in Table 2. WHO group III PH and WHO group V PH are frequently encountered in advanced lung disease.¹⁸ Numerous mechanisms for this association have been proposed, including destruction of the vascular bed by fibrosis, chronic hypoxemia, pulmonary veno-occlusive disease, compression of the vasculature by lymphadenopathy, and direct vascular involvement by the underlying disease process.¹⁹ The reasoning for exploring PH in advanced lung disease is sound because it provides important prognostic information and may also have implications for management. In IPF, Nadrous *et al.*²⁰ noted that echocardiographically diagnosed PH had a significant impact on survival. The median survival in patients with systolic pulmonary arterial pressure less than 35 mm Hg, between 36 and 50 mm Hg, and greater than 50 mm Hg was 4.8, 4.1, and 0.7 years, respectively. Lettieri *et al.*²¹ reiterated the point that pulmonary arterial pressure elevation is inversely related to survival. Because limited treatment options for IPF exist, PH remains an attractive target for therapy that requires further study in this population.

Methods of screening for PH in advanced lung disease vary. Echocardiography allows for structural visualization of the right and left heart in addition to the septum. Enlargement of the right heart with displacement of the septum in severe PH may be seen. Diastolic and systolic function of the left heart chambers can also be visualized with echocardiography.²² Additionally, the right ventricular systolic pressure can be estimated by the Bernoulli equation applied to the velocity of the tricuspid regurgitant jet.²³ However, the performance characteristics of echocardiography in evaluating PH in pulmonary fibrosis patients are unsatisfactory, limiting its use. To further define the pulmonary capillary wedge pressure and pulmonary artery

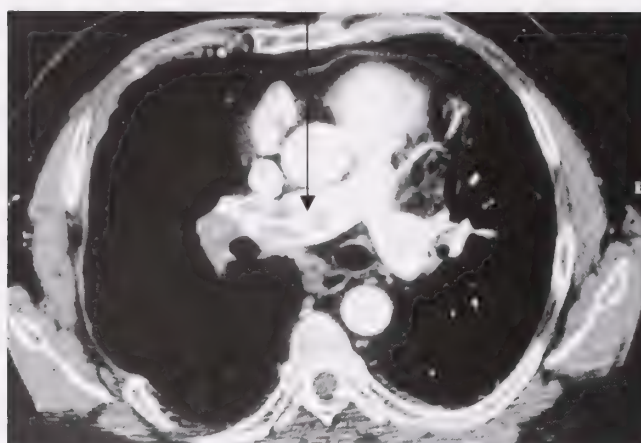


Fig 2. Large right main pulmonary arterial embolus.

Table 2. World Health Organization Classification of Pulmonary Hypertension.

Group I. Pulmonary arterial hypertension

- Idiopathic
- Familial
- Associated with the following:
 - Connective tissue disease
 - Congenital systemic-to-pulmonary shunts
 - Portal hypertension
 - Human immunodeficiency virus infection
 - Drugs and toxins
 - Other (thyroid disorders, glycogen storage disease, Gaucher's disease, hereditary hemorrhagic telangiectasia, hemoglobinopathies, myeloproliferative disorders, and splenectomy)
- Associated with significant venous or capillary involvement:
 - Pulmonary veno-occlusive disease
 - Pulmonary capillary hemangiomatosis
- Persistent pulmonary hypertension of the newborn

Group II. Pulmonary hypertension associated with left heart diseases

Group III. Pulmonary hypertension associated with respiratory diseases and/or hypoxemia (including chronic obstructive pulmonary disease)

Group IV. Pulmonary hypertension due to chronic thrombotic and/or embolic disease

Group V. Miscellaneous group (eg, sarcoidosis, histiocytosis X, and lymphangiomatosis)

NOTE: Revised clinical classification of pulmonary hypertension (Venice 2003).

pressure, catheterization of the right heart chambers is needed. Arcasoy *et al.*²⁴ reviewed a cohort of 374 lung transplant candidates with advanced lung disease. In only 44% of these patients was echocardiographic estimation of the right ventricular systolic pressure possible. Forty-eight percent of patients diagnosed with PH by echocardiography did not meet the diagnosis by cardiac catheterization. Consequently, echocardiography may be used as an adjunctive tool to screen patients with advanced lung disease for PH, whereas right heart catheterization remains the gold-standard diagnostic test.

Most studies attesting to the efficacy and safety of therapies for PH are from a select group of patients with WHO group I PAH. To date, there are 6 Food and Drug Administration–approved

therapies (Table 3) for PAH, including oral, inhaled, subcutaneous, and intravenous options.^{17,25} Because no current pharmacological therapy is effective for many forms of advanced lung disease, treating PH of advanced lung disease is appealing because even small improvements in functional status or quality of life would be worthwhile. Some centers are using therapy for PAH in patients with WHO group III or V PH with success. Collard *et al.*²⁶ enrolled a small cohort of patients with PH of IPF in an open-label study of sildenafil. Approximately 57% of these patients had a greater than 20% improvement in their 6-minute walk test distance, and the therapy was well tolerated. Inhaled prostacyclin has also been used as a therapy for PH of lung fibrosis.²⁷ These small series showing safety and

Table 3. *Therapy for Pulmonary Arterial Hypertension.*

Sildenafil (Revatio): oral
Bosentan (Tracleer): oral
Ambrisentan (Letairis): oral
Iloprost (Ventavis): inhaled
Treprostinil (Remodulin): subcutaneous or intravenous infusion
Epoprostenol (Flolan): intravenous infusion

efficacy underscore the need for more rigorous trials. Accordingly, patients with advanced lung disease and PH should be referred to centers at which consideration of enrollment into ongoing clinical trials is possible.

OBSTRUCTIVE SLEEP APNEA

Sleep hygiene may be deranged in advanced lung disease. Rates of death among patients with COPD who have hypercarbia and hypoxemia are high during sleep.²⁸ Obstructive sleep apnea (OSA) syndrome coexisting with COPD, the so-called overlap syndrome, is present in approximately 10% to 15% of patients with COPD.^{29,30} This likely increases the risks of PH, hypoxemia, and hypercarbia for patients with COPD.³⁰ Nocturnal treatment of COPD with noninvasive positive pressure ventilation (NIPPV) has been evaluated.³¹ Intuitively, treatment of the overlap syndrome with nocturnal NIPPV would have substantial benefit. Studies indicate that patients with severe CF may also have nocturnal desaturations. Instinctively, the consequences of this would be worsening PH and cor pulmonale. However, although daytime, nighttime, and exercise oxygen saturation reductions appear to correlate with PH in severe CF, the best indicator of concomitant PH in this patient population is daytime oxygen saturation reduction.³²

The coexistence of WHO group I PAH and OSA is not well defined. Several studies have tried to define the relationship, but criteria for the diagnosis of PAH have varied, and this makes the data difficult to interpret.³³ A recent literature review concluded that OSA was not common in IPAHA.³³ However, because OSA can lead to PH, it is often recommended that patients undergo a polysomnogram in their evaluation of PAH. If OSA is diagnosed, treatment with continuous positive airway pressure is likely beneficial but may not normalize the pulmonary pressures.³³

CARDIAC COMORBIDITIES

Cardiac causes of mortality in IPF, including ischemic heart disease (9.5%) and heart failure (14.4%), account for roughly one-quarter of deaths.² There is an increased prevalence of CAD in patients with fibrotic lung disease.³⁴ The risk of CAD in fibrotic lung disease increases further when multivessel disease is considered (odds ratio, 4.16; 95% confidence interval, 1.46–11.9). There are also data suggesting that significant CAD is more likely to accompany IPF than COPD.³⁵ However, the true prevalence of CAD in patients with COPD is unclear. In a cohort of 270 patients with COPD discharged from a university hospital, 10% were known to also have CAD.³⁶ The risks for CAD in IPF,³⁷ COPD, and sarcoidosis may also be aggravated by chronic or repetitive corticosteroid administration. Symptoms of concomitant cardiac disease can often masquerade as worsening lung disease. This underscores the need to be cognizant of multiple potential etiologies for worsening symptoms in patients with advanced lung disease.

ENDOCRINE COMORBIDITIES

Several hormones have been evaluated as attractive therapeutic targets for severe COPD.³⁸ With the offspring cohort of the Framingham Heart Study, the interrelation between diabetes mellitus (DM) and reduced pulmonary function was evaluated. Among patients who smoked and had DM, there was a reduction in the forced vital capacity of 251 mL in comparison with smokers without DM.³⁹ Whether this indicates that DM can cause an ongoing reduction in lung function is unclear. Furthermore, corticosteroid therapy for severe COPD may unmask glucose intolerance.⁴⁰ Likewise, DM has been evaluated as a link to IPF. Enomoto *et al.*⁴¹ noted that the odds ratio for IPF in diabetics was 4.06 (95% confidence interval, 1.80–9.15) in a case control study from Japan. Whether DM plays a role in the pathogenesis of advanced lung disease or whether it only accompanies it, perhaps worsened by corticosteroid use, is unclear. Nonetheless, treatment of hyperglycemia in patients with advanced lung disease is prudent. Finally, the relative hypogonadal state that can accompany COPD is likely due to hypoxemia. Although no compelling evidence is available that testosterone replacement therapy is necessary for the mild hypogonadal state in advanced lung disease,³⁸ some centers routinely evaluate and treat this deficiency.

HEMATOLOGIC COMORBIDITIES

Abnormalities of erythroid hematopoiesis are not uncommon in chronic lung disease. Chronic inflammation and hypoxemia are known causes of anemia and polycythemia, respectively. COPD is associated with anemia and, to a lesser degree, with polycythemia. In a review of stable patients with COPD, anemia was present in 17%, whereas polycythemia was present in 6%. Interestingly, in this review, anemia was associated with increased dyspnea, a shorter 6-minute walk test distance, and shorter survival in patients with COPD.⁴² Whether correcting the hemoglobin level changes these parameters is unknown.

Anemia has been described in sarcoidosis. However, this is felt to be due to bone marrow involvement with noncaseating granulomas and usually responds to steroid therapy.⁴³ Polycythemia in IPF is rare despite chronic hypoxemia.¹⁰ The mechanism of inefficient response to hypoxemia has been investigated but not fully defined.^{44,45} However, occasionally this may lead to anemia. Because erythropoietin levels are increased and no optimal level of hemoglobin is known, addressing this issue is of doubtful benefit. Regardless of the hemoglobin level, assessing for periods of hypoxemia with the implementation of supplemental oxygen as indicated is prudent.¹⁰

GASTROESOPHAGEAL REFLUX DISEASE

Gastroesophageal reflux affects approximately 10% to 20% of the population of the Western world.⁴⁶ This contrasts markedly with the prevalence of GERD in patients with many forms of advanced lung disease. The prevalence of GERD in IPF has been described as 87% to 94%. However, only 25% to 47% of patients in these studies noted symptoms of GERD.^{47,48} Raghu *et al.*⁴⁹ published a retrospective review of IPF patients treated solely with GERD therapy. This included proton pump inhibitor medications and fundoplication if needed. All 4 patients had improvement or stabilization of pulmonary function testing with adequate GERD therapy. Less than half of patients with IPF and concomitant GERD are symptomatic, and this underscores that this patient population should be screened and treated appropriately for GERD. Increased rates of GERD have also been associated with COPD. A recently published review of patients with COPD showed a 57% prevalence of GERD,⁵⁰ but interestingly, approximately

two-thirds of these patients were asymptomatic. With varying degrees of distal and proximal reflux and the low symptomatic rate, pH probe screening may be appropriate in defining patients with COPD in whom GERD should be treated.⁵⁰

PSYCHIATRIC COMORBIDITIES

Depression and anxiety have repeatedly been suggested as common comorbidities encountered in patients with advanced lung disease.⁵¹ Treatment of advanced lung disease with corticosteroids has also been associated with psychiatric side effects. Flaherty *et al.*³⁷ noted that in a cohort of patients treated with steroids for IPF, psychiatric symptoms were common. These included irritability in 61%, insomnia in 76%, depression in 37%, euphoria in 27%, psychosis in 10%, and fatigue in 32% of the patients. In a review of patients with PH, the diagnoses of depressive disorder and panic disorder were present in 15.9% and 10.4% of the cohort, respectively.⁵² These rates increased in conjunction with the New York Heart Association functional class, and this suggests that as PH progresses, so does the risk of psychiatric disorders. Untreated anxiety or depression is a worthy target for therapy in patients with advanced lung disease,⁵¹ and therefore screening for these is appropriate in the broad patient population. Lastly, with the known side effects of corticosteroids, iatrogenic psychiatric illness should be considered in patients on this form of therapy.

DECONDITIONING

Patients with advanced lung disease are often so dyspneic with minimal exertion that they fail to exercise.¹⁰ Superimposed anxiety due to breathlessness can further fuel this cycle of inactivity leading to severe deconditioning. Current recommendations for IPF and COPD include pulmonary rehabilitation in the comprehensive care of these diseases.^{10,53} Goals of pulmonary rehabilitation reach beyond increasing endurance and muscle tone. Social isolation, muscle wasting, depressed mood, and weight loss may all be positively affected by pulmonary rehabilitation. The benefits of pulmonary rehabilitation are not restricted to patients with parenchymal lung disease. In a recent study by Mereles *et al.*⁵⁴ of patients with PAH, exercise and respiratory training were safe and effective, increasing the 6-minute walk distance by a mean of 111 m. Thus, pulmonary rehabilitation should be considered in all patients with advanced

lung disease. Current recommendations underscore the importance of supplemental oxygen therapy if needed while patients are exercising.

OSTEOPOROSIS

Bone demineralization can be associated with many forms of advanced lung disease. In a recent study conducted to define the prevalence and predictors of osteoporosis in patients with diffuse parenchymal lung disease, 13% had osteoporosis, and 57% had osteopenia.⁵⁵ A low body mass index (BMI) and Hispanic race were found to predict low bone mineral density in this population. A low BMI in addition to corticosteroid administration has also been found to be a predictor of osteoporosis in a broad cohort of patients referred for lung transplantation.⁵⁶ Furthermore, severe WHO group I PAH has emerged as a risk for osteoporosis.⁵⁷ Not surprisingly, a low BMI also predicts osteoporosis in this population. Osteoporosis is thus strongly linked to severe lung disease, especially in patients with a low BMI. Whether corticosteroids play an independent or additive role is unclear. Because of the increased risk of osteoporosis after lung transplantation, symptomatic or severe osteoporosis remains a relative contraindication for this surgery.⁵⁸ Accordingly, these patients should be diligently screened and appropriately treated. This not only will

likely prevent complications of osteoporosis but also may augment transplant candidacy and outcomes.

MEDICATION EFFECTS

Corticosteroids are used for many forms of lung disease. Although no clear benefit has been ascribed to corticosteroids in IPF, corticosteroids in conjunction with other immunomodulating therapy may have benefit.^{10,59} Furthermore, corticosteroids have remained a mainstay of the medical armamentarium for COPD and sarcoidosis, although no clear mortality benefit has been proven.^{53,60} The use of corticosteroids has clear side effects that should be considered when decisions about this therapy are entertained. Known side effects of oral glucocorticoids include cataracts, peptic ulcer disease, myopathy, adrenal insufficiency, fluid retention, electrolyte imbalance, reduced immunity, and glucose intolerance. If the decision is made to institute corticosteroid therapy for advanced lung disease, a diligent watch for potential side effects is warranted.

SCREENING AND TREATING COMORBIDITIES

Although most advanced lung diseases are not part of a systemic process, care of the patient with advanced

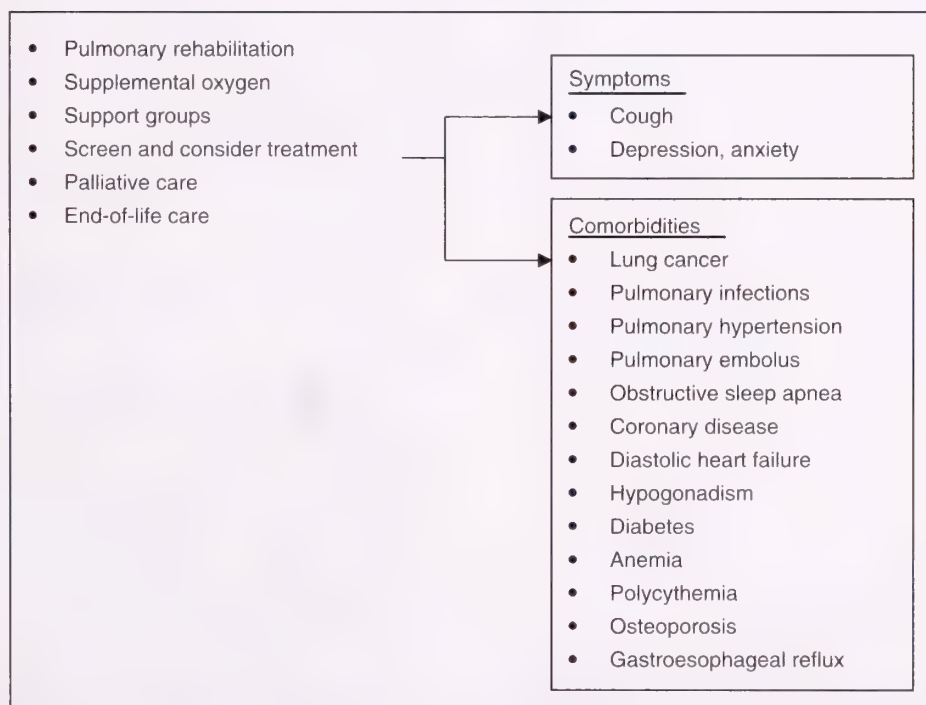


Fig 3. Adjunctive therapy for advanced lung disease.

lung disease can be viewed as caring for a systemic illness. No current guideline addresses all of the comorbidities that can be screened and potentially treated. However, with an understanding of the clinical continuum of each disease and the incidence and impact of common comorbid illnesses, a basis for the comprehensive medical care of these patients can be created and implemented (Figure 3). The management of the comorbid illnesses is similar to that of patients without advanced lung disease. In all cases, weighing the risks and benefits will help guide the implementation of possible therapies.

CONCLUSION

Advanced lung disease is an end result of numerous pulmonary processes that have defied therapy. Although differing in their etiology and course, they frequently converge in their manifestations, including disabling dyspnea and interceding comorbidities. The aggressive evaluation and treatment of comorbid illnesses may be more effective than any available disease-specific therapy. Therefore, acceptable goals of therapy include management of these comorbidities, other palliative measures, and limiting side effects of therapy. This allows for improvement in the quality of life for these patients and possibly an extension of life. For those who are appropriate lung transplant candidates, optimization of their pretransplant status facilitates posttransplant recovery and improves outcomes. As we anticipate novel effective therapies for the various advanced lung diseases, we should recognize that the evaluation and therapy of associated comorbidities are important and integral components of these patients' care.

ACKNOWLEDGMENT

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, the Department of Defense, or the US Government.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Hubbard R, Venn A, Lewis S, Britton J. Lung cancer and cryptogenic fibrosing alveolitis: a population-based cohort study. *Am J Respir Crit Care Med* 2002; 161: 5–8.
- Panos R. Clinical deterioration in patients with idiopathic pulmonary fibrosis: causes and assessment. *Am J Med* 1990; 88: 396–404.
- Aubry MC, Myers JL, Douglas WW, et al. Primary pulmonary carcinoma in patients with idiopathic pulmonary fibrosis. *Mayo Clin Proc* 2002; 77: 763–770.
- Bouros D, Hatzakis K, Labrakis H, Zeibecoglou K. Association of malignancy with diseases causing interstitial pulmonary changes. *Chest* 2002; 121: 1278–1289.
- Seersholm N, Vestbo J, Viskum K. Risk of malignant neoplasms in patients with pulmonary sarcoidosis. *Thorax* 1997; 52: 892–894.
- Wasswa-Kintu S, Gan WQ, Man SF, et al. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. *Thorax* 2005; 60: 570–575.
- Turner MC, Chen Y, Krewski D, et al. Chronic obstructive pulmonary disease is associated with lung cancer mortality in a prospective study of never smokers. *Am J Respir Crit Care Med* 2007; 176: 285–290.
- Saydain G, Islam A, Afessa B, et al. Outcome of patients with idiopathic pulmonary fibrosis admitted to the intensive care unit. *Am J Respir Crit Care Med* 2002; 166: 839–842.
- Nathan SD, Saleeb PG, Saft H, et al. Prevalence and impact of aspergillomas in patients with advanced sarcoidosis. *Am J Respir Crit Care Med* 2007; 175: A361.
- Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. *Am J Respir Crit Care Med* 2000; 161: 646–664.
- Magro CM, Allen J, Pope-Harman A, et al. The role of microvascular injury in the evolution of idiopathic pulmonary fibrosis. *Am J Clin Pathol* 2003; 119: 556–567.
- Nathan SD, Barnett SD, Urban BA, et al. Pulmonary embolism in idiopathic pulmonary fibrosis transplant recipients. *Chest* 2003; 123: 1758–1763.
- Kubo H, Nakayama K, Yanai M, et al. Anticoagulant therapy for idiopathic pulmonary fibrosis. *Chest* 2005; 128: 1475–1482.
- Fuster V, Steele PM, Edwards WD, et al. Primary pulmonary hypertension: natural history and the importance of thrombosis. *Circulation* 1984; 70: 580–587.
- Frank H, Mlczech J, Huber K, et al. The effect of anticoagulant therapy in primary and anorectic drug-induced pulmonary hypertension. *Chest* 1997; 112: 714–721.
- Rich S, Kaufmann E, Levy P. The effect of high doses of calcium-channel blockers on survival in primary pulmonary hypertension. *N Engl J Med* 1992; 327: 76–81.
- Badesch DB, Abman SH, Ahearn GS, et al. Medical therapy for pulmonary arterial hypertension: ACCP evidence-based clinical practice guidelines. *Chest* 2004; 126: 35S–62S.
- Simonneau G, Galiè N, Rubin LJ, et al. Clinical classification of pulmonary hypertension. *J Am Coll Cardiol* 2004; 43(suppl): 5S–12S.
- Nunes H, Humbert M, Capron F, et al. Pulmonary hypertension associated with sarcoidosis: mechanisms, haemodynamics and prognosis. *Thorax* 2006; 61: 68–74.

20. Nadrous HF, Pellicka PA, Krowka MJ, et al. The impact of pulmonary hypertension on survival in patients with idiopathic pulmonary fibrosis. *Chest* 2005; 128: 616S–617S.
21. Lettieri CJ, Nathan SD, Barnett SD, et al. Prevalence and outcomes of pulmonary arterial hypertension in advanced idiopathic pulmonary fibrosis. *Chest* 2006; 129: 746–752.
22. Bossone E, Bodini BD, Mazza A, Allegra L. Pulmonary arterial hypertension: the key role of echocardiography. *Chest* 2005; 127: 1836–1843.
23. Yock P. Noninvasive estimation of the right ventricular systolic pressure by Doppler ultrasound in patients with tricuspid regurgitation. *Circulation* 1984; 70: 657–662.
24. Arcasoy SM, Christie JD, Ferrari VA, et al. Echocardiographic assessment of pulmonary hypertension in patients with advanced lung disease. *Am J Respir Crit Care Med* 2003; 167: 735–740.
25. Galie N. Ambrisentan therapy for pulmonary arterial hypertension. *J Am Coll Cardiol* 2005; 46: 529–535.
26. Collard HR, Anstrom KJ, Schwarz MI, Zisman DA. Sildenafil improves walk distance in idiopathic pulmonary fibrosis. *Chest* 2007; 131: 897–899.
27. Olschewski H, Ghofrani HA, Walrath D, et al. Inhaled prostacyclin and iloprost in severe pulmonary hypertension secondary to lung fibrosis. *Am J Respir Crit Care Med* 1999; 160: 600–607.
28. McNicholas W. Nocturnal deaths among patients with chronic bronchitis and emphysema. *Br Med J (Clin Res Ed)* 1984; 289: 878.
29. McNicholas W. Impact of sleep in COPD. *Chest* 2000; 117: 48S–53S.
30. Chaouat A, Weitzenblum E, Krieger J, et al. Association of chronic obstructive pulmonary disease and sleep apnea syndrome. *Am J Respir Crit Care Med* 1995; 151: 82–86.
31. Wijkstra PJ, Lacasse Y, Guyatt GH, Goldstein RS. Nocturnal non-invasive positive pressure ventilation for stable chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2002; 2 CD002878.
32. Fraser KL, Tullis DE, Sasson Z, et al. Pulmonary hypertension and cardiac function in adult cystic fibrosis: role of hypoxemia. *Chest* 1999; 115: 1321–1328.
33. Atwood CW Jr, McCrory D, Garcia JG, et al. Pulmonary artery hypertension and sleep-disordered breathing: ACCP evidence-based clinical practice guidelines. *Chest* 2004; 126: 72S–77S.
34. Kizer JR, Zisman DA, Blumenthal NP, et al. Association between pulmonary fibrosis and coronary artery disease. *Arch Intern Med* 2004; 164: 551–556.
35. Reichner CA, Nathan SD, Barnett SD, et al. Prevalence and impact on outcomes of coronary artery disease in patients with IPF. *Am J Respir Crit Care Med* 2004; 169: A108.
36. Antonelli Incalzi R, Fuso L, De Rosa M, et al. Comorbidity contributes to predict mortality of patients with chronic obstructive pulmonary disease. *Eur Respir J* 1997; 10: 2794–2800.
37. Flaherty KR, Toews GB, Lynch JP III, et al. Steroids in idiopathic pulmonary fibrosis: a prospective assessment of adverse reactions, response to therapy, and survival. *Am J Med* 2001; 110: 278–282.
38. Creutzberg EC, Casaburi R. Endocrinological disturbances in chronic obstructive pulmonary disease. *Eur Respir J Suppl* 2003; 22: 76S–80S.
39. Walter RE, Beiser A, Givelber RJ, et al. Association between glycemic state and lung function: the Framingham Heart Study. *Am J Respir Crit Care Med* 2003; 167: 911–916.
40. Stanbury R, Graham E. Systemic corticosteroid therapy—side effects and their management. *Br J Ophthalmol* 1998; 82: 704–708.
41. Enomoto T, Usuki J, Azuma A, et al. Diabetes mellitus may increase risk for idiopathic pulmonary fibrosis. *Chest* 2003; 123: 2007–2011.
42. Cote C, Zilberberg MD, Mody SH, et al. Haemoglobin level and its clinical impact in a cohort of patients with COPD. *Eur Respir J* 2007; 29: 923–929.
43. Lower EE, Smith JT, Martelo OJ, Baughman RP. The anemia of sarcoidosis. *Sarcoidosis* 1988; 5: 51–55.
44. Tassiopoulos S, Kontos A, Konstantopoulos K, et al. Erythropoietic response to hypoxaemia in diffuse idiopathic pulmonary fibrosis, as opposed to chronic obstructive pulmonary disease. *Respir Med* 2001; 95: 471–475.
45. Tsantes A, Tassiopoulos S, Papadimitriou SI, et al. Suboptimal erythropoietic response to hypoxemia in idiopathic pulmonary fibrosis. *Chest* 2003; 124: 548–553.
46. Dent J, El-Serag HB, Wallander MA, Johansson S. Epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 2005; 54: 710–717.
47. Tobin RW, Pope CE II, Pellegrini CA, et al. Increased prevalence of gastroesophageal reflux in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1998; 158: 1804–1808.
48. Raghu G, Freudenberger TD, Yang S, et al. High prevalence of abnormal acid gastro-oesophageal reflux in idiopathic pulmonary fibrosis. *Eur Respir J* 2006; 27: 136–142.
49. Raghu G, Yang ST, Spada C, et al. Sole treatment of acid gastroesophageal reflux in idiopathic pulmonary fibrosis: a case series. *Chest* 2006; 129: 794–800.
50. Kempainen RR, Savik K, Whelan TP, et al. High prevalence of proximal and distal gastroesophageal reflux disease in advanced COPD. *Chest* 2007; 131: 1666–1671.
51. Wingate B, Hansen-Flaschen J. Anxiety and depression in advanced lung disease. *Clin Chest Med* 1997; 18: 495–505.
52. Löwe B, Gräfe K, Ufer C, et al. Anxiety and depression in patients with pulmonary hypertension. *Psychosom Med* 2004; 66: 831–836.
53. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176: 532–555.
54. Mereles D, Ehlken N, Kreuzer S, et al. Exercise and respiratory training improve exercise capacity and quality of life in patients with severe chronic pulmonary hypertension. *Circulation* 2006; 114: 1482–1489.
55. Caplan-Shaw CE, Arcasoy SM, Shane E, et al. Osteoporosis in diffuse parenchymal lung disease. *Chest* 2006; 129: 140–146.
56. Tschopp O, Boehler A, Speich R, et al. Osteoporosis before lung transplantation: association with low body

- mass index, but not with underlying disease. *Am J Transplant* 2002; 2: 167-172.
57. Tschopp O, Schmid C, Speich R, et al. Pretransplantation bone disease in patients with primary pulmonary hypertension. *Chest* 2006; 129: 1002-1008.
58. Orens JB, Estenne M, Arcasoy S, et al. International guidelines for the selection of lung transplant candidates: 2006 update—a consensus report from the Pulmonary Scientific Council of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant* 2006; 25: 745-755.
59. Raghu G, Depaso WJ, Cain K, et al. Azathioprine combined with prednisone in the treatment of idiopathic pulmonary fibrosis: a prospective double-blind, randomized, placebo-controlled clinical trial. *Am Rev Respir Dis* 1991; 144: 291-296.
60. Paramothayan S, Jones P. Corticosteroid therapy in pulmonary sarcoidosis: a systematic review. *JAMA* 2002; 287: 1301-1307.

Advanced Parenchymal Lung Disease: Quality of Life and Palliative Care

Christopher R. Gilbert, DO¹ and Cecilia M. Smith, DO²

¹Division of Pulmonary and Critical Care Medicine, Department of Medicine, Thomas Jefferson University Hospital, Philadelphia, PA

²Department of Medicine, The Reading Hospital and Medical Center, West Reading, PA

ABSTRACT

Advanced restrictive lung diseases remain a challenge for both the clinician and patient alike. Because there are few available treatment options that prolong survival for patients with diseases such as idiopathic pulmonary fibrosis, improvement in quality of life and palliation of significant symptoms become realistic treatment goals. Several validated instruments that assess quality of life and health-related quality of life have demonstrated the dramatic impact that lung disease has on patients. Quality-of-life assessments of patients with interstitial lung disease have commonly cited respiratory complaints as problematic, but other distressing symptoms often not addressed include fear, social isolation, anxiety, and depression. Not only do respiratory symptoms limit this patient population, but the awareness of decreased independence and ability for social participation also has an impact on the quality of life. Some patients describe a deepened spiritual well-being during their disease process; however, many patients' mental health suffers with experiences of fear, worry, anxiety, and panic. Many patients express desire for more attention to end-of-life issues from their physicians. Fears of worsening symptoms and suffocation exist with an expressed desire by most to die peacefully with symptom control. Interventions to improve quality of life are largely directed at symptom control. Pharmacologic and nonpharmacologic interventions have been helpful in relieving dyspnea. Studies have demonstrated that the use of supplemental oxygen in the face of

advancing hypoxemia can have both positive and negative effects on quality of life. Patients using nasal prongs describe feelings of self-consciousness, embarrassment, and social withdrawal. Pulmonary rehabilitation is recommended, with some studies noting increased quality-of-life scores and decreased sensations of dyspnea. Sleep deprivation and poor sleep quality also have a negative impact on quality of life. Recognition and correction of nocturnal hypoxemia and other sleep disturbances should enhance quality of life in patients with restrictive lung disease; however, there is currently no evidence to support this claim. End-of-life care needs more attention by clinicians in the decision-making and preparatory phase. Physicians need to maintain their focus on quality-of-life issues as medical management shifts from curative therapies to comfort management therapies. Palliative care and hospice appear to be underused in patients with advanced diseases other than cancer. Because the only curative option for some end-stage restrictive lung diseases is lung transplantation, if transplantation is not an option, palliation of symptoms and hospice care may offer patients and families the opportunity to die with dignity and comfort. *Mt Sinai J Med* 76:63–70, 2009. © 2009 Mount Sinai School of Medicine

Key Words: advanced interstitial lung disease, dyspnea, end-of-life care, hospice, palliative care, quality of life.

Many issues surround the care of patients diagnosed with end-stage lung disease. Extensive research exists on patients with obstructive lung disease [ie, chronic obstructive pulmonary disease (COPD)] with respect to quality of life (QOL) and palliative care measures; however, there is limited literature regarding these issues in patients with advanced lung disease from restrictive defects [ie, idiopathic pulmonary fibrosis (IPF)].

In 2000, a consensus statement from the American Thoracic Society explained the often progressive and terminal nature of IPF. Median

Address Correspondence to:

Cecilia M. Smith, D.O.

Department of Medicine
The Reading Hospital and
Medical Center

West Reading, PA

Email: smithcm@readinghospital.org

survival rates of 2 to 4 years after diagnosis along with 5-year survival rates ranging from 30% to 50% have been described.¹ Therapy for IPF is controversial, as no randomized trials have definitively proven benefits of medical treatment in reversing the disease process or prolonging survival.¹⁻⁴ The only option for increased survival appears to be lung transplantation; however, not all patients are eligible candidates.⁵⁻⁷

These sobering statistics require a realistic look at what should be the most important goal of treatment for patients diagnosed with IPF and other end-stage fibrotic lung diseases. If no significant treatment options exist for curing their disease, then management of symptoms and maximization of QOL become goals on which physicians managing these patients need to focus. The American College of Chest Physicians (ACCP) strongly supports the position that palliative and end-of-life care should be an integral part of the management of patients with chronically progressive pulmonary disease.⁸

QUALITY OF LIFE

The World Health Organization defines health as "not merely the absence of disease or infirmity, but as a concept that incorporates notions of well-being or wellness in all areas of life (physical, mental, emotional, social, spiritual)."⁹ Alternatively, Flanagan¹⁰ divided QOL into 5 basic categories:

- Physical and material well-being.
- Relations with other people.
- Social, community, and civic activities.
- Personal development and fulfillment.
- Recreation.

These categories have been used in attempts to objectively assess QOL in patients living with chronic diseases. Health-related quality of life (HRQL) has also been recognized as an important tool in determining the impact that chronic disease states may have on patients' perceptions of their own health. These assessments can be used to help gauge the impact that chronic lung disease may have on patients' lives.^{11,12}

There are 4 frequently used questionnaires that can be used to assess QOL in chronic lung disease. These include the following: the World Health Organization 100-Item Instrument, the Medical Outcomes Study Short-Form 36-Item Instrument (SF-36), the Chronic Respiratory Questionnaire, and the St. George's Respiratory Questionnaire (SGRQ).¹³

Common symptoms cited as affecting QOL in patients with chronic interstitial lung diseases (ILDs)

include progressive breathlessness, fatigue, dry cough,¹¹ decreased activity levels, sleep disturbances, loss of independence, decreased libido, social isolation, fear, anxiety, decreased mobility,¹³ and depression.¹⁴

Dyspnea or progressive breathlessness appears to be the most common and distressing symptom affecting QOL in patients with IPF.^{11-13,15-17} During their evaluation of the SF-36 questionnaire in patients with IPF, Martinez *et al.*¹⁵ demonstrated that dyspnea is the most important factor limiting patients' HRQL. They also noted similar findings in patients with end-stage COPD.

The most common complaint, besides dyspnea and cough, that has a negative impact on patients' lives is fatigue. Patients make significant changes to their lives in order to function throughout the day. A number of patients noted the comfort that they took from knowing the love and support that they would receive from their family, but at the same time, they were concerned about the physical, financial, and emotional burden being placed on their family members.¹³

Jastrzebski *et al.*¹⁷ evaluated QOL measures in patients that were wait-listed for lung transplantation using the SF-36 questionnaire. Patients with IPF, when compared to a COPD population, demonstrated significant limitations in physical functioning, social functioning, and emotional limitations. After 1 year on the transplant list, there was significant worsening in QOL, specifically with respect to physical limitations.

A limited number of studies have attempted to determine the best method to appropriately evaluate QOL in patients with ILD.^{12,15,18,19} Studies have tried to use validated general and disease-specific questionnaires to determine impairments in comparison with the general population. There have also been attempts to use more objective measurements such as pulmonary function testing; high-resolution computed tomography, arterial blood gases, and 6-minute walk test distance to determine the impact that IPF may have on patients.^{12,17,20-22}

Using the SF-36, Martinez *et al.*¹⁵ concluded that patients with IPF demonstrate significant impairments in physical and psychological functioning. Pulmonary function testing, specifically the forced vital capacity and forced expiratory volume over 1 second, correlated with physical functioning and general health perceptions. A negative correlation was demonstrated with respect to the pH and components such as physical functioning, emotional role, vitality, and social functioning. They also described dyspnea as the most important factor influencing QOL, noting similar findings in patients with COPD. IPF patients

waiting for a lung transplant with a higher partial pressure of oxygen in arterial blood and forced expiratory volume over 1 second were in better general health, and this indicates that objective measures may be useful in predicting QOL.¹⁶

The following study by Swigris *et al.*¹³ offers a number of insights into realms of QOL of which we as physicians may not be aware but in which we may be able to intervene in order to improve QOL in patients suffering from IPF.

Twenty patients were invited to participate in focus groups along with individual interview sessions to discuss the impact that IPF had on their lives. Problems commonly cited included low energy, sleep disturbances, and exhaustion. This also manifested itself in problems with decreased sexual activity, decreased independence (including activities of daily living), and decreased social participation. Patients were often dissatisfied with pharmacological treatment options, noting that the side effects were "worse than the disease itself."¹³

Some patients feared traveling into public places for fear of catching a respiratory illness that might lead to their death. Others commented that they spent a great deal of time "planning ahead," noting that they had to be aware of how they would travel on different excursions so as to not "overdo it." Many voiced concerns over finances and the strain that they may place on their families. Patients stated that they did not want to exhaust family savings on their medical care, noting that "in terms of using up my finances, continuing living is a real concern of mine."¹³

Some commented that their diagnosis led to improved spiritual well-being, but their mental health generally suffered as they often reported fear, worry, anxiety, and panic regarding their future mortality. Many patients were aware of their expected disease course and expressed desire for more attention to end-of-life issues from their physicians. They also admitted fear of their worsening symptoms and fear of death as a "conscious suffocation." They hoped that when their death was imminent, they could be assured that their symptoms would be controlled and that their passing would be peaceful.¹³

In summary, Swigris *et al.*¹³ was able to demonstrate that there are tools available to evaluate the significant adverse effects that IPF has on QOL; however, these questionnaires also miss important points that have dramatic impacts on patients' QOL.

INTERVENTIONS TO IMPROVE QUALITY OF LIFE

As mentioned before, no pharmacological therapy has ever been proven to reverse or alter the inflammatory process of IPF,¹ and some treatment regimens may decrease QOL as a result of increased side effects.^{23,24} Few studies have specifically evaluated treatments intended to improve QOL in patients with IPF. Some of the more commonly used treatments are aimed at symptom control or palliation, including oxygen, narcotics, benzodiazepines, and pulmonary rehabilitation. Other avenues that have been explored include the impact of treatment on sleep-related breathing disorders (SRBDs) and pulmonary hypertension in patients with advanced restrictive lung disease.

Nonpharmacological interventions to help relieve dyspnea are used in hospice patients but have not been evaluated in patients with advanced restrictive lung disease. Guided imagery, which both is easy to teach and has no side effects, has been noted to produce improvement in QOL in patients with COPD. Energy conservation, pursed-lip breathing, and diaphragmatic training are widely accepted as ways of alleviating dyspnea, but there is no significant scientific evidence to validate their effectiveness.²⁵

As mentioned previously, dyspnea and breathlessness are common symptoms of patients with end-stage lung disease.^{11-13,15-17} Opioids are commonly administered to help with the symptomatic treatment of dyspnea in patients with end-stage restrictive lung disease. However, this practice is not supported by significant data. A randomized, double-blinded, placebo-controlled trial of mostly patients with obstructive lung disease demonstrated that sustained-release oral morphine (20 mg) helped with the management of refractory dyspnea.²⁶ Jennings *et al.*²⁷ reviewed the available evidence for the use of opioids in managing dyspnea from any cause and concluded that oral and parental opioids are appropriate for palliation. However, only 4 of the 18 studies included patients with restrictive lung disease for a total of 14 of 293 patients. They also concluded that is no beneficial effect of nebulized opioids in comparison with nebulized saline in relieving dyspnea. Some healthcare workers remain hesitant to provide narcotic medications for fear of inducing respiratory depression and hastening death.²⁸ Contrary to this belief is evidence that in patients with cancer-related dyspnea, the use of morphine may increase survival by diminishing physical exhaustion and psychological distress.^{29,30}

Smoller *et al.*³¹ reviewed the treatment of panic and anxiety in lung disease patients with respect to the relief of anxiety and improvement of QOL. Panic can cause patients to experience sensations of depersonalization, derealization, and fears of going crazy. These behaviors affect patients' QOL by altering their behavior in anticipation of future attacks.³² There is a lack of scientifically validated tools for assessing panic and anxiety in patients with end-stage lung disease, and most assessments appear to be performed subjectively. Benzodiazepines are commonly used, as these are effective anxiolytic agents. However, incorrect use of these medications can lead to respiratory depression and hypercapnia, especially in patients with concomitant sleep-disordered breathing. Lorazepam has been noted to induce less respiratory depression than other medications in the benzodiazepine class. Tricyclic antidepressants and selective serotonin reuptake inhibitor medications are effective for the treatment of panic disorders. Amitriptyline, a tricyclic antidepressant, has demonstrated an ability to improve mood, increase functional status, and decrease respiratory symptoms in patients with asthma. The results for COPD have been equivocal, and other lung diseases have not been extensively studied. Selective serotonin reuptake inhibitors have shown an antidyspneic and anxiolytic effect in patients with COPD.³¹

Cough has been described as both a presenting and complicating symptom associated with essentially all of the ILDs. Cough is a common symptom of many other pulmonary and nonpulmonary diseases. Therefore, the ACCP recommends that common etiologies, such as asthma, gastroesophageal reflux, and postnasal drip, be explored first before ILD is diagnosed as the cause of cough.³³

If the etiologic basis of the cough is IPF, it represents a debilitating symptom that may affect HRQL. Treatment modalities for chronic cough have always been directed at identifying and eliminating the underlying cause; however, if a cause is not found, symptomatic treatment is often attempted.³⁴ Common antitussive therapies such as codeine and dextromethorphan have not been studied in ILD. A recent observational phase II trial using thalidomide (100–400 mg/day) noted complete resolution of cough symptoms in 10 of 11 patients.³⁴ Corticosteroid use has been generally recommended for treating cough in IPF, sarcoidosis, and hypersensitivity pneumonitis. In patients with IPF, corticosteroids may lead to symptomatic improvement, but they have not demonstrated the ability to prolong survival or QOL.^{33,35}

Oxygen therapy is a common therapeutic intervention offered to patients with lung disease,

especially if there is documented evidence of hypoxemia. One study noted that approximately 65% of patients enrolled in hospice programs for end-stage lung disease or lung cancer were treated with oxygen, mostly with the intention of symptom control.²⁵ Patients with IPF often demonstrate hypoxemia, if not at rest then very commonly with minimal exercise and sleep.¹ Oxygen administration appears beneficial in nonhypoxemic COPD patients, as it provides an increased exercise capacity,³⁶ however, no studies in nonhypoxemic, restrictive disease have been reported. A large retrospective study by Douglas *et al.*²⁴ demonstrated that oxygen therapy in IPF was associated with decreased survival, yet it was unclear if oxygen therapy was simply a marker for more severe disease or if there truly was an adverse effect of oxygen. However, a multivariate analysis demonstrated that oxygen therapy was not associated with shortened survival, but at the same time, no survival benefit was noted in patients using oxygen therapy. No effects on QOL were evaluated in this study. Oxygen therapy is not without its problems as it appears that it may worsen QOL in patients with advanced lung disease. Some patients note self-consciousness, embarrassment, and social withdrawal with the use of nasal prongs in public.²¹ Tomioka *et al.*³⁷ demonstrated that IPF patients in Japan using long-term oxygen therapy had significant detriments in the physical and psychosocial domains of the SF-36 questionnaire over a 1-year follow-up period. Another study using the SF-36 and SGRQ noted worse scores in a low-income COPD population that used oxygen therapy.³⁸

There is also an obvious cost element associated with oxygen at home. An analysis of monthly energy costs for various utility companies was performed for a 400-W oxygen concentrator. Monthly costs ranged from \$51.36 in southern California to \$11.16 in Dalton, GA. Many patients may qualify for financial rebates; however, families may still suffer financial hardship in order to continue their oxygen therapy.³⁹

Despite data suggesting that QOL may worsen with oxygen therapy, many patients do not let their oxygen therapy significantly limit their activities, including travel. Visiting family or enjoying vacation destinations continues to be an option available for patients on supplemental oxygen, which most consider a positive enhancement to QOL. With coordination, scheduling, and added cost in some situations, trips can be made by car, bus, train, boat, and airplane.⁴⁰

Policies about bringing oxygen vary by bus, rail, or cruise line. Plane trips require similar coordination and advance planning to arrange

oxygen in flight but also require a note from a physician stating the liter flow per minute so that the airline can calculate how much oxygen the patient will use throughout the flight. A Web site sponsored by the Airline Oxygen Council of America (www.airlineoxygenCouncil.org) offers detailed information for travelers and healthcare providers.⁴⁰ There is a high risk of in-flight hypobaric hypoxemia for patients with chronic restrictive lung disease.⁴¹ However, with supplemental oxygen, oxygen saturations can be maintained at acceptable levels.^{42–44}

Referral to pulmonary rehabilitation has been recommended as a standard of care for chronic lung diseases according to recent guidelines from the ACCP and American Association of Cardiovascular and Pulmonary Rehabilitation. Although limited studies have used pulmonary rehabilitation in non-COPD populations, the guidelines provide a grade 1B recommendation that pulmonary rehabilitation is beneficial for patients with chronic lung diseases other than COPD.⁴⁵ In patients with IPF, pulmonary rehabilitation has demonstrated improvement in QOL at 8 weeks along with some sustained improvement at 1 year.⁴⁶ Jastrzebski *et al.*⁴⁷ were able to demonstrate that pulmonary rehabilitation caused a decreased sensation of dyspnea along with improvement in all QOL markers according to the SF-36 questionnaire. There do not appear to be any data available to definitively conclude that pulmonary rehabilitation decreases hospitalizations or mortality in patients with IPF. As noted, there is little scientific evidence to support specific rehabilitation regimens, but expert opinion suggests that treatment strategies should be tailored to specific diseases and patients.⁴⁵

Pulmonary hypertension occurs in up to 50% of patients with IPF that are undergoing evaluation for lung transplantation.⁴⁸ Pulmonary vasodilators have demonstrated their effect in improving the 6-minute walk test in patients with pulmonary hypertension and IPF. Although no formal testing for QOL has been evaluated, patients have been noted to have a meaningful improvement in functional status.⁴⁹ One may be able to extrapolate that QOL may improve if exercise tolerance and functional capacity improve.

The true incidence of SRBDs in patients with restrictive lung disease is unknown because of a lack of sufficient literature. Studies have demonstrated that nocturnal hypoxemia is common in patients with IPF.³⁹ Using the SF-36, Clark *et al.*⁵⁰ was able to demonstrate that decreased scores in physical functioning, social functioning, and emotional domains were associated with decreased overnight oxygen saturation and an abnormal sleep study. Perez-Padilla *et al.*⁵¹ found that patients with

severe ILD had decreased REM sleep, along with increased stage 1 and fragmented sleep. These findings may lend support to the notion that poor sleep quality may have a significant impact on these patients' lives. Polysomnographic studies in patients with IPF have demonstrated significant sleep architecture alterations, even in the absence of obstructive sleep apnea. This is thought to be related to the increased ventilatory drive common in IPF patients.^{51,52} Another study evaluating SRBDs in patients with IPF noted that 61% complained of exhaustion and daytime fatigue, thought to be related to sleep apnea, which had a negative impact on their QOL.⁵² Although there appears to be preliminary evidence to support screening for SRBDs in patients with IPF, there are no studies confirming that treatment of SRBDs improves QOL or survival in this patient population.

Home care is another option that can provide services such as patient assessment, education, counseling, and symptom management, along with home medical equipment. It is often assumed that home healthcare models are cost-effective; however, significant data to support this theory are lacking. The dyspnea that many patients with end-stage lung disease experience may be difficult for family members and some healthcare workers to manage at home, and they therefore may need referral to an acute care facility.⁵³ Some QOL issues appear to improve with home healthcare in the terminally ill,⁵⁴ however, there are no substantial data to support this idea in patients with ILD. COPD patients noted improvement in the SGRQ after 4 months of home healthcare.⁵⁵ Dyspnea, independence in activities of daily living, and HRQL do not improve with the use of home healthcare. There is an increase in family caregiver satisfaction when their ill family member is enrolled in home health services.⁵²

END-OF-LIFE CARE

After establishing the diagnosis of an end-stage lung disease such as IPF, the clinician, by education and preparation by assisting the patient and family in facing the patient's limited life expectancy. However, the course of any one patient's illness can vary, and this makes it difficult to predict when a patient will die.⁵⁶ One study found that IPF patients were shocked by the condition of other patients with more severe IPF, indicating that patients may not be fully aware of their own limited life expectancy. This same study also noted that patients view a diagnosis of cancer as much worse than IPF.¹⁹ Nonetheless,

patients with IPF are known to have a limited survival of 3 to 5 years.¹ This knowledge alone should lead the clinician to prepare patients for the course of their disease. A clinician's own comfort level, experience, and education in matters of end-of-life care may be a barrier or a resource to a patient facing this serious illness.⁵⁷ There are a number of educational resources available to physicians and nurses on the Internet. A recent ACCP position statement regarding palliative care in patients offers an appendix full of references for healthcare professionals, family, and patients.⁸

Patients with ILD may look for guidance, education, and support from their physician as their disease advances.⁸ Physicians can assist the patient in decision-making and preparation for end-of-life considerations. This includes medical decisions that can have an impact on the experience for the patient during the dying process. Symptoms of pain, depression, and anxiety may affect patient preferences for medical care. Depressive symptoms appear related to patient preferences for resuscitation.⁵⁸ In COPD patients, depressed patients refused resuscitation at a greater rate than patients without depression. COPD patients who improved with treatment for depression were 5 times more likely to change their preference toward requesting resuscitation.⁵⁹ Effects of depression and end-of-life treatment preferences have not been studied in ILDs. These findings have potential implications for medical providers, should be taken into account. The depressed state of the patient when discussing end-of-life care. However, one must remember that the role of certain aggressive medical treatments in advanced ILD may be limited. In 1 study of IPF patients admitted to the intensive care unit for acute respiratory failure, only 2 of 15 patients survived to be discharged from the hospital. In this study, the benefit of mechanical ventilation in advanced disease was poor, and acute respiratory failure did not improve with mechanical ventilation.⁶⁰

Patients look to their healthcare providers to initiate such a discussion, ideally in the outpatient setting, well before end-stage disease is apparent. End-of-life issues, while difficult to discuss for some patients, are best explored at a noncritical time. This allows the patient to consider his or her own end-of-life wishes and discuss these wishes with his or her spouse, significant other, family, and friends. End-of-life care, implemented on the basis of a patient's considered wishes, is an intervention that may ease tension, fear, and stress for both the patient and the family as the disease advances. The patient is treated with dignity and respect throughout the process, and with the patient's wishes followed, families are often relieved of the responsibility of making difficult decisions.

Focusing on patients' QOL during medical treatment should not cease when it is decided that aggressive medical therapy is no longer appropriate. As the disease continues to progress in the face of maximum medical therapy, a shift in goals of therapy and alternative management plans should be considered, discussed, and implemented. Issues and concerns identified as affecting QOL for patients with chronic lung disease at an earlier stage of the disease may become magnified as symptoms worsen and limitations increase. As mentioned previously, patients with IPF have yearned for more attention to end-of-life issues from the medical community. Patients have expressed insecurity about the dying process, fear of a lack of control of symptoms toward the end of their lives, and fear of experiencing conscious suffocation.¹³

Palliative care and hospice care provide patients with chronic lung disease treatment options that attempt to control their symptoms and improve QOL. Palliative care and hospice services have long been associated with support for terminal cancer patients. A survey of acute care and cardiorespiratory nurses revealed the misperception that palliative care is available only for patients with malignancies.⁶¹ Hospice services are currently available to patients regardless of diagnosis, yet they appear to be underused for patients dying with advanced stages of nonmalignant lung diseases. Hospice seeks neither to prolong life nor to hasten death. Rather, its purpose is to maximize comfort, dignity, and QOL for the dying, along with offering help to surviving family members with bereavement services.⁶²

Appropriate hospice referrals generally require that physicians believe their patient has a life expectancy of less than 6 months; however, patients with advanced lung disease often have unpredictable courses. Guidelines for referral are available from Medicare and private health insurers. A review by Abrahm and Hansen-Flaschen⁶² notes the questionable validity of these guidelines in the modern era and instead offers a common-sense approach for referral to hospice. Their criteria include progressive lung disease in which the patient is at high risk of death from an intercurrent illness, adequate medical therapy without significant improvement, potential benefit of hospice services for palliation of distressing symptoms, and patient acceptance of death with a wish to avoid unnecessary suffering.

Because IPF is amenable to lung transplantation, treatment modalities diverge in patients with advanced disease between hospice, which provides comfort and ameliorates symptoms and candidacy for transplantation. In the appropriate candidate, lung transplantation can be a therapeutic option; however,

there is a high incidence of death while patients wait for transplantation.⁶³ When transplantation is not an option, patients could potentially benefit from hospice, which offers patients the opportunity to die with dignity and comfort.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. *Am J Respir Crit Care Med* 2000; 161: 646–664.
- Raghu G, Brown KK, Bradford WZ, et al. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2004; 350: 125–133.
- Johnson MA, Kwan S, Snell NJ, et al. Randomised controlled trial comparing prednisolone alone with cyclophosphamide and low dose prednisolone in combination in cryptogenic fibrosing alveolitis. *Thorax* 1989; 44: 280–288.
- Raghu G, Depaso WJ, Cain K, et al. Azathioprine combined with prednisone in the treatment of idiopathic pulmonary fibrosis: a prospective double-blind, randomized, placebo-controlled clinical trial. *Am Rev Respir Dis* 1991; 144: 291–296.
- Unilateral lung transplantation for pulmonary fibrosis. Toronto Lung Transplant Group. *N Engl J Med* 1986; 314: 1140–1145.
- Thabut G, Mal H, Castier Y, et al. Survival benefit of lung transplantation for patients with idiopathic pulmonary fibrosis. *J Thorac Cardiovasc Surg* 2003; 126: 469–475.
- Selman M. From anti-inflammatory drugs through antifibrotic agents to lung transplantation: a long road of research, clinical attempts, and failures in the treatment of idiopathic pulmonary fibrosis. *Chest* 2002; 122: 759–761.
- Selecky PA, Eliasson CAH, Hall RI, et al. Palliative and end-of-life care for patients with cardiopulmonary diseases: American College of Chest Physicians position statement. *Chest* 2005; 128: 3599–3610.
- Constitution of the World Health Organization. *Chronicle World Health Org* 1947; 1: 29–43.
- Flanagan JC. A research approach to improving our quality of life. *Am Psychol* 1978; 33: 138–147.
- Swigris JJ, Kuschner WG, Jacobs SS, et al. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005; 60: 588–594.
- Baddini Martinez JA, Martinez TY, Lovetro Galhardo FP, de Castro Pereira CA. Dyspnea scales as a measure of health-related quality of life in patients with idiopathic pulmonary fibrosis. *Med Sci Monit* 2002; 8: CR405–CR410.
- Swigris J, Stewart A, Gould M, Wilson S. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005; 3: 61.
- Vries JD, Kessels BLJ, Drent M. Quality of life of idiopathic pulmonary fibrosis patients. *Eur Respir J* 2001; 17: 954–961.
- Martinez TY, Pereira CAC, dos Santos ML, et al. Evaluation of the short-form 36-item questionnaire to measure health-related quality of life in patients with idiopathic pulmonary fibrosis. *Chest* 2000; 117: 1627–1632.
- Jastrzebski D, Kozielski J, Banas A, et al. Quality of life during one-year observation of patients with idiopathic pulmonary fibrosis awaiting lung transplantation. *J Physiol Pharmacol* 2005; 56(suppl 4): 99–105.
- Swigris JJ, Gould MK, Wilson SR. Health-related quality of life among patients with idiopathic pulmonary fibrosis. *Chest* 2005; 127: 284–294.
- Nishiyama O, Taniguchi H, Kondoh Y, et al. Health-related quality of life in patients with idiopathic pulmonary fibrosis. What is the main contributing factor? *Respir Med* 2005; 99: 408–414.
- De Vries J, Seebregts A, Drent M. Assessing health status and quality of life in idiopathic pulmonary fibrosis: which measure should be used? *Respir Med* 2000; 94: 273–278.
- De Vries J, Drent M. Quality of life and health status in interstitial lung diseases. *Curr Opin Pulm Med* 2006; 12: 354–358.
- Reynolds HY. Diagnostic and management strategies for diffuse interstitial lung disease. *Chest* 1998; 113: 192–202.
- Martinez FJ, Safrin S, Weycker D, et al. The clinical course of patients with idiopathic pulmonary fibrosis. *Ann Intern Med* 2005; 142(pt 1): 963–967.
- Douglas WW, Ryu JH, Swensen SJ, et al. Colchicine versus prednisone in the treatment of idiopathic pulmonary fibrosis: a randomized prospective study. *Am J Respir Crit Care Med* 1998; 158: 220–225.
- Douglas WW, Ryu JH, Schroeder DR. Idiopathic pulmonary fibrosis. Impact of oxygen and colchicine, prednisone, or no therapy on survival. *Am J Respir Crit Care Med* 2000; 161: 1172–1178.
- Webb M, Moody LE, Mason LA. Dyspnea assessment and management in hospice patients with pulmonary disorders. *Am J Hosp Palliat Med* 2000; 17: 259–264.
- Abernethy AP, Currow DC, Frith P, et al. Randomised, double blind, placebo controlled crossover trial of sustained release morphine for the management of refractory dyspnoea. *BMJ* 2003; 327: 523–528.
- Jennings AL, Davies AN, Higgins JPT, et al. A systematic review of the use of opioids in the management of dyspnoea. *Thorax* 2002; 57: 939–944.
- Ajemian I. Palliative management of dyspnea. *J Palliat Care* 1991; 7: 44–45.
- Dudgeon DJ, Rosenthal S. Management of dyspnea and cough in patients with cancer. *Pain Palliat Care* 1996; 10: 157–171.
- Bruera E, MacEachern T, Ripamonti C, Hanson J. Subcutaneous morphine for dyspnea in cancer patients. *Ann Intern Med* 1993; 119: 906–907.
- Smoller JW, Pollack MH, Otto MW, et al. Panic anxiety, dyspnea, and respiratory disease. Theoretical and clinical considerations. *Am J Respir Crit Care Med* 1996; 154: 6–17.
- Periyakoil VS, Skultety K, Sheikh J. Panic, anxiety, and chronic dyspnea. *J Palliat Med* 2005; 8: 453–459.

33. Brown KK. Chronic cough due to chronic interstitial pulmonary diseases: ACCP evidence-based clinical practice guidelines. *Chest* 2006; 129(suppl): 180S–185S.
34. Horton MR, Danoff SK, Lechtzin N. Thalidomide inhibits the intractable cough of idiopathic pulmonary fibrosis. *Thorax* 2008; 63: 749.
35. Madison JM, Irwin RS. Chronic cough in adults with interstitial lung disease. *Curr Opin Pulm Med* 2005; 11: 412–416.
36. Emtner M, Porszasz J, Burns M, et al. Benefits of supplemental oxygen in exercise training in nonhypoxemic chronic obstructive pulmonary disease patients. *Am J Respir Crit Care Med* 2003; 168: 1034–1042.
37. Tomioka H, Imanaka K, Hashimoto K, Iwasaki H. Health-related quality of life in patients with idiopathic pulmonary fibrosis—cross-sectional and longitudinal study. *Intern Med* 2007; 46: 1533–1542.
38. Sant'Anna CA, Stelmach R, Zanetti Feltrin MI, et al. Evaluation of health-related quality of life in low-income patients with COPD receiving long-term oxygen therapy. *Chest* 2003; 123: 136–141.
39. Reisfield GM, Wilson GR. The cost of breathing: an economic analysis of the patient cost of home oxygen therapy. *Am J Hosp Palliat Med* 2004; 21: 348–352.
40. Traveling with oxygen. Available at: http://www.yourlunghealth.org/healthy_living/articles/traveling/index.cfm. Accessed June 2008.
41. Gendreau MA, DeJohn C. Responding to medical events during commercial airline flights. *N Engl J Med* 2002; 346: 1067–1073.
42. Martin SE, Bradley JM, Buick JB, et al. Flight assessment in patients with respiratory disease: hypoxic challenge testing vs. predictive equations. *QJM* 2007; 100: 361–367.
43. Christensen CC, Ryg MS, Kare Refvem O, Henning Skjonsberg O. Effect of hypobaric hypoxia on blood gases in patients with restrictive lung disease. *Eur Respir J* 2002; 20: 300–305.
44. Seccombe LM, Kelly PT, Wong CK, et al. Effect of simulated commercial flight on oxygenation in patients with interstitial lung disease and chronic obstructive pulmonary disease. *Thorax* 2004; 59: 966–970.
45. Ries AL, Bauldoff GS, Carlin BW, et al. Pulmonary rehabilitation: joint ACCP/AACVPR evidence-based clinical practice guidelines. *Chest* 2007; 131(suppl): 4S–42S.
46. Naji NA, Connor MC, Donnelly SC, McDonnell TJ. Effectiveness of pulmonary rehabilitation in restrictive lung disease. *J Cardiopulm Rehabil* 2006; 26: 237–243.
47. Jastrzebski D, Gumola A, Gawlik R, Kozielski J. Dyspnea and quality of life in patients with pulmonary fibrosis after six weeks of respiratory rehabilitation. *J Physiol Pharmacol* 2006; 57(suppl 4): 139–148.
48. Nadrous HF, Pelikka PA, Krowka MJ, et al. Pulmonary hypertension in patients with idiopathic pulmonary fibrosis. *Chest* 2005; 128: 2393–2399.
49. Collard HR, Anstrom KJ, Schwarz MI, Zisman DA. Sildenafil improves walk distance in idiopathic pulmonary fibrosis. *Chest* 2007; 131: 897–899.
50. Clark M, Cooper B, Singh S, et al. A survey of nocturnal hypoxaemia and health related quality of life in patients with cryptogenic fibrosing alveolitis. *Thorax* 2001; 56: 482–486.
51. Perez-Padilla R, West P, Lertzman M, Kryger MH. Breathing during sleep in patients with interstitial lung disease. *Am Rev Respir Dis* 1985; 132: 224–229.
52. Statement on home care for patients with respiratory disorders. *Am J Respir Crit Care Med* 2005; 171: 1443–1464.
53. Hansen-Flaschen J. Chronic obstructive pulmonary disease: the last year of life. *Respir Care* 2004; 49: 90–97 discussion 97–98.
54. Hughes SL, Weaver FM, Giobbie-Hurder A, et al. Effectiveness of team-managed home-based primary care: a randomized multicenter trial. *JAMA* 2000; 284: 2877–2885.
55. Bourbeau J, Julien M, Maltais F, et al. Reduction of hospital utilization in patients with chronic obstructive pulmonary disease: a disease-specific self-management intervention. *Arch Intern Med* 2003; 163: 585–591.
56. Fox E, Landrum-McNiff K, Zhong Z, et al. Evaluation of prognostic criteria for determining hospice eligibility in patients with advanced lung, heart, or liver disease. *JAMA* 1999; 282: 1638–1645.
57. Emanuel LL, Barry MJ, Stoeckle JD, et al. Advance directives for medical care—a case for greater use. *N Engl J Med* 1991; 324: 889–895.
58. Stapleton RD, Nielsen EL, Engelberg RA, et al. Association of depression and life-sustaining treatment preferences in patients with COPD. *Chest* 2005; 127: 328–334.
59. Rosenfeld KE, Wenger NS, Phillips RS, et al. Factors associated with change in resuscitation preference of seriously ill patients. The SUPPORT Investigators. *Arch Intern Med* 1996; 156: 1558–1564.
60. Blivet S, Philit F, Sab JM, et al. Outcome of patients with idiopathic pulmonary fibrosis admitted to the ICU for respiratory failure. *Chest* 2001; 120: 209–212.
61. Davidson P, Introna K, Daly J, et al. Cardiorespiratory nurses' perceptions of palliative care in nonmalignant disease: data for the development of clinical practice. *Am J Crit Care* 2003; 12: 47–53.
62. Abraham JL, Hansen-Flaschen J. Hospice care for patients with advanced lung disease. *Chest* 2002; 121: 220–229.
63. Nathan SD. Lung transplantation: disease-specific considerations for referral. *Chest* 2005; 127: 1006–1016.

ETHICS

Fistula Experience in Niger: How We Overcame an Ethical Dilemma

Cedric K. Olivera, MD, MS, Charles J. Ascher-Walsh, MD, MS
and Nada Gligorov, PhD

Mount Sinai School of Medicine, New York, NY

A vesicovaginal fistula is an abnormal connection between the bladder and vagina. The extent of the urologic injury depends on several factors in the obstructed labor course. One can have a vesicouterine, vesicocervical, vesicovaginal, or vesicourethral fistula due to an absent urethra. The type of pelvis, size of the fetus, and type of labor dysfunction will determine where the fistula is located. In Niger, most of the fistulas that we encountered were distal, involving the vesicourethral junction or even part or all of the urethra. There is no consensus for staging fistulas; however, some practitioners use location, size, or recurrence. Because of the prevalence of injuries to the urethra, we classified women as having partial or complete traumatic absence of the posterior urethra (TAPU). Rectovaginal fistulas, abnormal connections between the rectum and vagina, were also treated by the surgical team.

Obstructed labor is the most common cause of vesicovaginal fistulas in the developing world. The sizes of fistulas vary and may not reflect the extent of the damage to the surrounding tissue. Twenty to thirty percent of women continue to leak urine after successful repair of a fistula because of stress incontinence or because of the small bladder capacity after the use of adjacent tissue to repair the defect. Many women suffer from stress incontinence, likely because of a combination of decreased bladder size and increased pressure once the defect is repaired.

Neuropathic bladder dysfunction can also result from ischemic injury to the nerves that run through the bladder base and trigone. As the bladder is pulled up and compressed against the pubic symphysis, it is very common for the upper third of the urethra to be compromised and sloughed off as part of the fistula. This leads to a shortened urethra. If the compression occurs more distally, the fistula can occur as an isolated defect at the urethra, leading to TAPU. Complete TAPU is seen in 5% of women with obstetric fistulas; partial TAPU is seen in 30%. Most of these women require urethral reconstruction to attain continence. Other known risk factors for fistula formation are a rudimentary infrastructure, poor access of the patients to midwifery or emergency obstetrical care, malnutrition, and marriage at a young age (under 20) before the pelvis has had time to completely develop.

Reviewing the fistula data from John Hopkins in 1958, Everett and Mattingly¹ were able to demonstrate that 65% of the fistulas were caused by either gynecologic surgery (40%) or obstetrics (24%). Some of the sequelae associated with an obstructed labor pattern include fetal injury (95% stillborn), fistula formation, amenorrhea, vaginal stenosis, neurological injury associated with foot drop,² and social injury, including divorce, isolation, and depression.

Fistula formation and repair have been well documented in history. Marion Sims, a renowned gynecologic surgeon, operated on enslaved African women with fistulas. One woman in particular, Anarcha, required multiple surgical attempts until continence was finally achieved with silver wire. Debate continues concerning whether or not his continued attempts at closure with multiple previous failures were in the best interest of his patient Anarcha from an ethical standpoint. Where should one draw the line between innovation and abuse?

In the developed world, the most common cause of fistula formation is benign gynecologic

Address Correspondence to:

Cedric K. Olivera

Mount Sinai School of Medicine
New York, NY

Email: rosamond.rhodes@
mssm.edu

surgery. These fistulas tend to be small, and many may even heal on their own. Surgery is a viable option for fistulas that do not heal on their own. Spontaneous healing is possible because the women in developed countries have improved nutrition, health insurance, and access to a surgeon who can actually perform the procedure.

WOMEN AND FISTULAS IN NIGER, WEST AFRICA

Niger is a country with an area of 490,000 square miles, about 3 times the size of California, of which two thirds is desert. It has a population of just over 11 million. The illiteracy rate is 85%. Fifty-six percent of the population is Hausa, and 22% is Djerma. French is the official language, although Hausa and Djerma are 2 of the more popular dialects spoken. Ninety-five percent of the country is Islamic.

Unlike the typical case in the United States, for women in places such as Niger, fistulas develop from labor;³ these women are often in labor for days and often end up with a dead fetus. They come to the hospital as quickly as possible, but for many of them, this means an 11-hour donkey cart ride to the nearest hospital. For most of the women who choose to wait it out and deliver at home, the labor course is grossly obstructed, and the fetal head is wedged in an immature pelvis incapable of allowing its passage for hours. The net result is ischemia, which is followed by necrosis and ultimately fistula formation. For women with an obstructed labor course who make it to a hospital before significant hemorrhaging occurs, the usual case scenario is delivery of a dead fetus via cesarean section.

Unfortunately, very few of the doctors in the country are trained to repair a fistula. Because of the shortage of trained physicians who can operate on these women, the women with this condition are forced to live a life of solitude, reeking of urine and ostracized by their husbands and community. Typically, they seek refuge in a community of their peers, a fistula community of women with a similar plight. In a 1966 publication, E. Catherine Hamlin describes the plight of women with fistulas:

Constantly in pain, incontinent of urine or feces, bearing the heavy burden of sadness of discovering their child stillborn, ashamed of a rank personal offensiveness, abandoned therefore by their husbands, outcasts

of society, employable only in the fields, they live, they exist, without friends and without hope.

One woman described her predicament:

We went out looking to be healed, but they said we were all whores. This sickness caught me when I was only a young girl. I sat confused at home until I heard the good news. I said, my word, mother! Give me the money. I will go to Jankwano in os! I will go down there and see the new Europeans. By this time tomorrow I will have arrived in the operating theatre. I will remain there at Dr. Karshima's place. Doctor Karshima I thank you! May God bless you! You have sewn up the leaking woman. My husband threw me out because I was leaking. If this sickness catches you they'll carry you out and throw you away too.

Is an African expression for a woman with a fistula.

OUR MISSION

The International Organization for Women's Development is an organization in New York that sponsors trips to Niger by teams of surgeons who can repair fistulas. The first time that a fistula is repaired is the best time for a cure because the tissue is most responsive to healing. Unfortunately for the women in Niger, many of them have had surgeries in the past before competent fistula surgeons could perform the repair. Equally unfortunate is the fact that even when the fistulas are repaired by competent surgeons, after each repair, the success rate for cure diminishes. The anticipated cure rate for a primary fistula repair is in the range of 80% to 90%, but the cure rate falls precipitously after an initial failure.

When a woman is suspected of having a fistula, a careful examination of her history and a physical examination are vitally important. In Niger, where resources are limited, advanced radiologic technology is not readily available; methylene blue is the chief diagnostic tool. Once the diagnosis is made, attempts are made to bolster the patient's nutrition so that healing can be facilitated. The keys to a successful repair are adequate surgical exposure, wide mobilization, a tension-free closure in as many layers as possible, and prolonged bladder drainage.⁴⁻⁶

The women in Niger who are afflicted with fistulas and are serviced by the International Organization for Women's Development live in the courtyard of the National Hospital in Naimey. Their mattresses reek of urine. They patiently wait for someone to come and fix their problem surgically. The mean time for surgical correction is 3.5 years.

The way that we deal with multiple prior surgical repair failures truly becomes an ethical dilemma. To learn more about the population in Niger, a questionnaire was developed and distributed. Fifty-eight women who underwent surgical repair for vesicovaginal fistulas at the National Hospital Fistula Center in Niamey, Niger, from September 2005 to January 2006 were asked to complete the questionnaire. Each questionnaire was administered orally in either Hausa or Djerma by a fluent interviewer. Data were analyzed with the SPSS statistics program. The results were as follows:

- Mean age at marriage: 15.6 years (range, 10–22 years; standard deviation, 2.46 years).
- Mean age at first pregnancy: 17.31 years (range, 8–25 years; standard deviation, 3.36 years).
- Mean age at time of survey: 26.14 years, (range, 16–44; standard deviation, 7.5 years).
- Percentage of the study population considered to be married: 62%.
- Percentage of women who began their labor at home and were primarily cared for by a family member: 94.8%.
- Mean length of labor: 2.61 days.
- Mean length of labor before further assistance was sought: 1.61 days.
 - 91.4% had gone to a healthcare center by the time of delivery.
 - 91.4% of the women gave birth to stillborn infants.
- Percentage of infants dead by 2 days postpartum: 96.6%.
 - 63.8% were dead via vaginal delivery.
 - 13.8% were dead via forceps-assisted vaginal delivery.
 - 22.4% were dead via cesarean section.

It was also learned that that there were significant barriers to the women seeking healthcare. In 2002, there were only 6 surgeons in Niger who could repair fistulas, and only 3 had actually received training in this area. We found that the mean travel time to a hospital was 11.52 hours, and although 56.9% arrived by bush taxi, 6.8% arrived by donkey cart.

Oral consent for surgery was obtained preoperatively by translators fluent in Hausa and Djerma and in the patient's own dialect. The information that was provided included an explanation of risks, benefits, and alternative treatment options. We learned that postoperatively 44.8% of the patients stated that no one had explained the procedure to them. Only 31.3% of the patients felt that they had received an explanation and had actually understood what was being said to them; this shows the problem of

obtaining informed consent in a patient population in which the illiteracy rate is 85%.

ONE PATIENT

K.M. is a 32-year-old para 4. Like so many other women, she was forced to live in a state of constant leaking of urine after her fistula developed. This 32-year-old woman could not be sexually active, and she wanted to be. She was ostracized from her community.

K.M. presented to us at the city hospital in Niamey to have her fistula evaluated. She had already had 6 previous vesicovaginal fistula repairs, which had proven to be unsuccessful. Because of repeated fistula repair and scarring of the vagina, she had a foreshortened vagina that basically ended in a blind pouch and a urethra that was severely scarred. K.M. wanted to undergo another surgery, hoping that perhaps we could restore some normalcy to her life.

K.M. presented us with difficult ethical dilemmas. Should we attempt a seventh repair of her fistula? Should we tell her that we could not do anything to help her because after the fourth fistula repair the success rate drops to around 17%? Should we offer her a urinary diversion, a procedure carrying a 30% mortality rate in which the ureters would be reimplanted into the terminal ileum? Could we adequately inform this patient of the options? Could we obtain her true informed consent for any procedure?

We had a paucity of clinical outcome data to guide our decision-making process in this scenario. Our intentions were good, but without data, any attempt at treatment would be somewhat experimental.

We understood that patients should be educated and informed about what treatment options are available. The risks, benefits, and alternatives to those treatment options should also be discussed. An agreement should be reached between the patient and her physician for the treatment plan. These are basic ethical assumptions for treating patients both clinically and as subjects of research: the desire to first do no harm and to respect the patient's autonomy. These principles are the moral foundation for the informed-consent process.

After considering all of the issues and ethical difficulties related to communicating with this patient, we decided to comply with K.M.'s request, respect her autonomy, and perform her seventh fistula repair operation. Immediately after the operation, she was dry, and the operation was a success. However, several weeks later, the patient presented

with leakage again. Although the treatment of this particular patient could be viewed as a failure, we (the surgical team), in conjunction with the patient, dealt with an ethical dilemma in the best way that we knew how. In retrospect, we ask ourselves if we should have managed the patient differently. The answer is no because we made an attempt to obtain informed consent, respect the patients autonomy, and first do no harm.

It has been said that

To have danced with the fistula patient is to be submerged in a unique community of caring, which is bound together by an almost indescribable knowledge of the nature of suffering. To have danced with the fistula patients is to recall why you went into medicine in the first place, and to mourn what we in the West have lost, despite our advanced health care, affluent economies and technical expertise.

That truly was our experience.

DISCLOSURES

There is nothing to report.

REFERENCES

1. Everett HS, Mattingly RF. Vesicovaginal fistula. *Am J Obstet Gynecol* 1956; 72: 712–724.
2. Elkins TE. Surgery for the obstetric vesico vaginal fistula: a review of 100 operations in 82 patients. *Am J Obstet Gynecol* 1994; 170: 108–1118.
3. Nafiou I, Idrissa A, Ghaichatou AK, et al. Obstetric vesico-vaginal fistulas at the National Hospital of Niamey, Niger. *Int J Gynaecol Obstet* 2007; 99(suppl 1): S71–S74.
4. Arrowsmith SD. Genito-urinary reconstruction in obstetric fistulas. *J Urol* 1994; 152: 403–406.
5. Waaldijk K. Classification of VVF, Babbar Ruga Fistula Hospital.
6. Margolis T, Mercer LJ. Vesico vaginal fistulas. *Obstet Gynecol Surv* 1994; 49: 840–847.

SPECIAL FEATURE

Acetaminophen Hepatotoxicity: Current Management

Hui-Hui Tan, MBBS, MRCP,^{1,2} Charissa Y. Chang, MD,¹ and Paul Martin, MD³

¹Division of Liver Diseases, Recanati-Miller Transplantation Institute, Mount Sinai School of Medicine, New York, NY

²Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore

³Division of Hepatology, Centre for Liver Diseases, University of Miami School of Medicine, Miami, FL

Acetaminophen [*N*-acetyl-*p*-aminophenol (APAP)], either singly or as a component of a drug cocktail, is a frequently used over-the-counter and prescription medication worldwide with analgesic and antipyretic properties, and it has a reasonable safety profile when consumed in therapeutic doses. In contrast, supratherapeutic doses (intentionally or inadvertently ingested) were recognized to cause severe hepatotoxicity and even fatal hepatic failure as early as the 1960s.¹ Acute liver failure is rapid deterioration in liver function (>26 weeks' duration) leading to altered mentation (encephalopathy), typically with coagulopathy [usually international normalized ratio (INR) ≥ 1.5], in a patient without preexisting liver disease.² Although acute viral hepatitis A, B, and E are the main causes of acute liver failure in developing countries, APAP toxicity is now the commonest cause of acute liver failure in the United States and United Kingdom.^{3–7} APAP protein adducts can be detected in more than 17% of indeterminate cases of acute liver failure, and this suggests that APAP hepatotoxicity can be implicated in a substantial minority of cryptogenic causes of acute liver failure.^{8,9} Hence, APAP toxicity should be specifically excluded in any patient with unexplained acute hepatic dysfunction.

DRUG PHARMACOKINETICS AND PROFILE

APAP is rapidly and completely absorbed from the gastrointestinal tract. The serum concentration usually peaks within 2 hours of therapeutic dose ingestion and is reached by 4 hours of ingestion of an overdose in short-acting preparations.¹⁰ However, peak serum levels may be affected by simultaneous ingestion with drugs that alter gastric emptying or be delayed beyond 4 hours after an overdose of extended-release preparations.^{11–13}

The half-life of APAP is 2 to 4 hours for all preparations. However, sustained-release preparations may result in delayed elimination phases due to prolonged tablet dissolution or absorption.^{12,14}

The therapeutic limit is 15 mg/kg per dose in children and 1 g per dose in adults every 4 to 6 hours up to a maximum daily dose of 80 mg/kg in children or 4 g in adults. The use of therapeutic doses of APAP is not associated with increased risk of hepatotoxicity in patients with chronic liver disease.¹⁵

The therapeutic serum concentration of APAP ranges from 10 to 20 $\mu\text{g/mL}$. However, the diagnosis of APAP-induced hepatotoxicity should be entertained in a patient with a history of chronic, excessive APAP ingestion and elevated liver enzymes, regardless of measured serum levels. Toxicity is unlikely to occur with a single dose of less than 150 mg/kg in children or less than 7.5 g in an adult.¹⁶ Conversely, it is likely to occur when amounts greater than 250 mg/kg in children or more than 12 g in adults are ingested over a 24-hour period.^{17,18}

Address Correspondence to:

Hui-Hui Tan
Department of
Gastroenterology and
Hepatology,
Singapore General Hospital,
Singapore
Email: mhyhui@gmail.com

DRUG METABOLISM

At the usual therapeutic doses, the majority (up to 90%) of APAP is metabolized in the liver to form sulfate and glucuronide conjugates. These conjugates are then excreted in the urine¹⁰ (Figure 1). The remainder of the drug is either excreted unchanged in the urine or metabolized via the cytochrome P450 pathway into a hepatotoxic metabolite, *N*-acetyl-*p*-benzoquinone imine (NAPQI).^{19,20} At therapeutic doses, the small amount of NAPQI that is produced forms a benign conjugate with glutathione, which is then excreted in the urine. However, at toxic doses of APAP, sulfation and glucuronidation pathways are rapidly saturated, and more of the drug is then metabolized via the cytochrome P450 pathway, forming NAPQI.

NAPQI formation and liver injury are also influenced by other factors apart from excessive APAP ingestion. Induction of the cytochrome P450 pathway by other drugs (eg, chronic alcohol ingestion), decreased host capacity for glucuronidation, and depletion of glutathione stores (eg, malnutrition and chronic alcohol ingestion) also increase the formation of NAPQI.

DIAGNOSIS

The presentation of APAP poisoning is variable and ranges from initial asymptomatic mild hepatic dysfunction to frank hepatic failure. It should also be remembered that initial clinical manifestations do not necessarily predict subsequent hepatotoxicity.^{21,22} In addition to recognizing that a potentially lethal dose

has been ingested, the physician should perform a thorough evaluation, eliciting whether other drugs or agents (eg, anticonvulsants or chronic alcohol ingestion that may induce the cytochrome P450 system) are involved and determining the time of ingestion, intent (suicidal or accidental), pattern of ingestion, and existence of comorbid conditions (eg, malnutrition) that may predispose to hepatic injury or NAPQI formation. A recent retrospective study of 6386 patients with APAP-induced liver injury suggested that older age, female gender, and the concomitant use of sympathetic stimulants (but not statins) were associated with increased odds for medical support–required events.²³ However, further studies will be required to confirm these findings.

The differential diagnosis of acute hepatic dysfunction includes alcoholic hepatitis, other drug-induced or toxin-induced hepatitis, viral hepatitis, Reye's syndrome, and ischemic hepatitis. Interestingly, marked serum total bilirubin elevations (>10 µg/mL) may cause false-positive serum APAP levels.^{24–26} However, the actual threshold of hyperbilirubinemia for this phenomenon to occur is unclear.

PREDICTING HEPATIC INJURY

There is a poor correlation between the self-reported quantity of APAP ingested and detected serum APAP levels.^{27,28} The Rumack-Matthew nomogram predicts the risk of hepatotoxicity and indication for *N*-acetylcysteine (NAC) therapy after a single, acute ingestion of APAP overdose.²⁹ In Rumack and Matthew's original retrospective study, 64

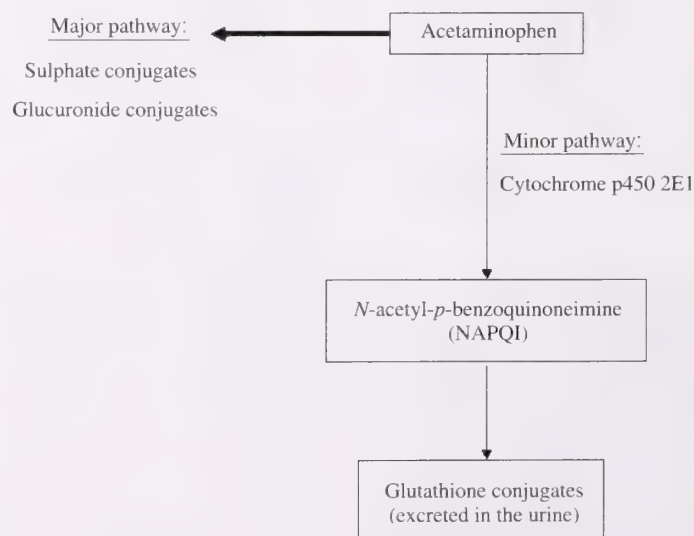


Fig 1. Acetaminophen metabolism at therapeutic doses.

patients with APAP toxicity were evaluated without antidote therapy. Serum APAP levels correlated with severe hepatotoxicity (arbitrarily defined as aspartate aminotransferase levels >1000 IU/L). Serum levels >200 $\mu\text{g/mL}$ drawn at 4 hours or serum levels >6.25 $\mu\text{g/mL}$ drawn at 24 hours from the time of ingestion predicted severe hepatotoxicity (60% probability) or mortality (5% probability). In the United States, this "probability line" was lowered by 25% as additional "built-in safety" from the time when the US Food and Drug Administration first approved an early study for the use of NAC as an antidote to acetaminophen poisoning.³⁰ The main disadvantage of this nomogram is that its use is validated only in the setting of an acute, single ingestion with a clear history. Unfortunately, this is not the only pattern of liver injury in patients who present with overdose. Also, severe hepatotoxicity has been recognized to occur even when serum APAP levels fall below the probability line. However, little has been described about the latter group of patients. In a study of 2540 patients, this incidence was 3.2%.^{30,31}

UNINTENTIONAL OVERDOSE WITH N-ACETYL-P-AMINOPHENOL

Alcohol use was initially thought to be an important risk factor for APAP hepatotoxicity, and the accidental nature of APAP toxicity in alcoholics was termed *therapeutic misadventure*.^{32,33} This was attributed to their risk for an induced-enzyme state, which makes them more susceptible to liver injury. Several conditions have been reported that may predispose a patient to this risk, such as fasting, malnutrition, and, most commonly, alcohol ingestion. However, this remains controversial. Data have been conflicting as the fasted or malnourished state, which may decrease glutathione levels, may also decrease CYP2E1 levels and alcohol acts as both a substrate and an inducer of CYP2E1. Results from clinical studies and reports have also borne out this controversy and it is debatable if APAP ingestion at purely therapeutic doses is more harmful in alcoholics.^{18,30,34–37} From a physiological standpoint, acute APAP hepatotoxicity at unequivocally therapeutic doses is extremely unlikely, despite case reports of therapeutic misadventure. In most such cases, the drug is found to have been consumed in excessive doses, and analyzing the various case reports has been difficult as data are often incomplete.^{38,39} Hence, the term *unintentional overdose* is more appropriate.

Although suicidal overdose of APAP is the most frequent form of liver injury in the United

Kingdom, unintentional cases occur more commonly in the United States.⁴⁰ Unlike suicidal APAP overdose, overdoses that are unintentional tend to be associated with ingestion over several days, a lack of awareness of hepatotoxicity risk on the part of the patient, pain that is usually chronic, depression, and polysubstance abuse.^{4,5,41} Late presentation is characteristic in unintentional overdose as medical attention is sought only after symptoms have developed, and this results in a poorer prognosis. The US Acute Liver Failure Study Group has identified salient clinical features in this group of patients.⁴ Often, there is repeated dosing when pain relief is not adequate, the use of a narcotic combination, the combined ingestion of more than 1 acetaminophen-containing product, or a combination of these. Although total ingestion is usually consumed over an average of 3 days, the median dose consumed for this group of patients is 34 g (versus 20 g for suicidal intent). However, once acute liver failure occurs, the outcomes of the two types of overdose (intentional and unintentional) are similar.⁴²

TOOLS FOR PROGNOSTICATION IN ACUTE LIVER FAILURE (TABLE 1)

King's College Criteria

The King's College criteria are the most frequently used for prognostication in APAP toxicity.^{18,43,44} These criteria have good specificity but poorer sensitivity. In the original report by O'Grady et al.⁴⁵ published in 1989, the clinical and laboratory features of 588 patients with acute liver failure between 1973 to 1985 were retrospectively reviewed to identify factors predictive of a poor prognosis on univariate and multivariate analyses. These were then tested retrospectively in another 175 patients admitted to the authors' unit from 1986 to 1987, and this led to the construction of a model for the selection of patients for liver transplantation. Poor prognostic factors in APAP toxicity are pH <7.3 after fluid resuscitation or serum creatinine >3.3 mg/dL, prothrombin time >100 seconds (or >1.8 times above the control or INR <6.4), and the presence of grade III-IV encephalopathy. The main limitation of this model is that therapeutic interventions (eg, the correction of coagulopathy) may affect key values, negating their predictive value.

Arterial Lactate Levels

Incorporating arterial lactate levels into the King's College criteria has been proposed, and they are used

Table 1. Predictors of Poor Prognosis in APAP-Induced Acute Liver Failure.

Criteria	Parameter/Predictor
King's College ¹⁵	pH <7.3 (after fluid resuscitation) or PT >100 seconds (or PT >1.8 times control or INR >6.5), serum creatinine >300 μ mol/L (>3.3 mg/dL), and grade III-IV encephalopathy
Arterial lactate ⁴⁶	Arterial lactate >3.0 mmol/L after fluid resuscitation
Coagulation factors ⁴⁷	Factor V level <10% (at presentation) and factor VIII/factor V ratio >30 (within 3 days of presentation)
APACHE II ⁴⁸	APACHE II score >15 (at presentation)
Serum phosphate ^{49*}	Serum phosphate level >1.2 mmol/L (48–96 hours after APAP ingestion)
MELD ⁵⁴	MELD >33 (on day of onset of encephalopathy)

Abbreviations: APACHE II, Acute Physiology and Chronic Health Evaluation II; APAP, *N*-acetyl-p-aminophenol; INR, international normalized ratio; MELD, Model for End-Stage Liver Disease; PT, prothrombin time.

* Conflicting data have been reported (see the main text).

in some parts of Europe to assist with prognostication of relevant patients. In Bernal et al.'s⁴⁰ retrospective cohort study, arterial lactate >3.5 mmol/L before fluid resuscitation or arterial lactate >3.0 mmol/L after fluid resuscitation was found to predict poor outcomes. However, lactate levels before fluid resuscitation had poorer sensitivity and poorer accuracy than the King's College criteria. Conversely, arterial lactate levels after fluid resuscitation had equivalent sensitivity, equivalent time to predicting poor prognosis, and greater specificity and greater accuracy in comparison with the King's College criteria. These findings were also further validated in a prospective study.⁴⁶

Coagulation Factor Levels

A study from King's College Hospital of 27 patients from 1989 to 1990 with acute liver failure identified a factor V level <10% on presentation and a factor VIII/factor V ratio >30 within 3 days of presentation as predictive of a poor outcome. These values predicted death or the need for a liver transplant with an accuracy of 95%. Among the 27 patients

studied, 22 patients had acute liver failure from APAP poisoning; of these, 11 survived, whereas the rest died or had a liver transplant.⁴⁷ All survivors had factor VIII/factor V ratios <30, and the median factor V level was 11.8% on presentation among survivors (versus a median factor V level of 5.1% among those with poor outcomes). The main disadvantages of using coagulation factors are a lack of ready access to factor level assays and the fact that coagulopathy may be temporarily corrected by the administration of blood products.

Acute Physiology and Chronic Health Evaluation II Scoring

Acute Physiology and Chronic Health Evaluation (APACHE) II scoring upon admission has been applied to APAP hepatotoxicity to predict outcome.⁴⁸ In a study of 102 patients with acetaminophen poisoning, APACHE II scoring was more sensitive and more specific than the King's College criteria. Unfortunately, the tool is comparatively long and cumbersome with several variables to score. However, it may predict a poor prognosis at the time of initial presentation.

Serum Phosphate Levels

A prospective study published in 2002 enrolled 125 patients with severe APAP-induced hepatotoxicity.⁴⁹ Serial serum phosphate levels were measured. A level >1.2 mmol/L between 48 and 96 hours after APAP ingestion predicted a poor chance of survival in the absence of a liver transplant. However, these findings were not confirmed in other studies.^{50,51} It has recently been suggested that hypophosphatemia in acute liver failure patients, in general, may be due to renal phosphate wasting instead of internal redistribution of phosphate alone.⁵² Additional studies are needed to validate serum phosphate levels as a prognostic tool in APAP hepatotoxicity.

Model for End-Stage Liver Disease Scores

In a more recent study,⁵³ the Model for End-Stage Liver Disease (MELD) scoring system (available at <http://www.unos.org/resources/meldpeldcalculator.asp>) was prospectively applied to 460 patients with APAP hepatotoxicity from a single center between the years 1999 and 2004. Hepatic encephalopathy developed in 63 of 142 patients with a MELD score >18 at 48 to 72 hours after ingestion (positive predictive value: 44%) versus 2 of 182 patients with a MELD score <18 (negative predictive value: 99%). Among

124 patients with acute liver failure, a threshold MELD score of 33 on the day of onset of encephalopathy as a predictor of death or need for liver transplant had a sensitivity of 60%, a specificity of 69%, a positive predictive value of 65%, and a negative predictive value of 63%. Thus, this study found that MELD scores may be useful as a predictor of fulminant hepatic failure in patients with APAP poisoning. However, as a predictor of death from fulminant hepatic failure, the discriminative power of MELD scoring was not superior to the King's College criteria or INR alone.

Several other small studies have evaluated other parameters and indices^{54–56} in an effort to find a better prognostic tool for APAP hepatotoxicity. However, none have yet been developed that surpass the utility of the King's College criteria.

MANAGEMENT

N-Acetylcysteine

The antidote for APAP hepatotoxicity is NAC. It is recommended in all patients in whom the quantities of APAP ingested, serum drug levels, or rising aminotransferases indicate a risk of hepatotoxicity.² Its use is also suggested in patients with acute liver failure when APAP ingestion is possible or even when knowledge of circumstances surrounding admission is inadequate.²

When administered early (within 8 hours of APAP ingestion), NAC limits the accumulation of NAPQI by directly binding to it, increasing glutathione stores and increasing sulfate conjugation.⁵⁷ No deaths have been reported in larger studies in which NAC was administered within 10 hours of APAP ingestion, regardless of serum levels.^{31,58,59} Additional benefits when it is administered later may include anti-inflammatory and anti-oxidative effects.^{60,61} These theoretical benefits are less well characterized and less effective than its primary effects when NAC is administered earlier following APAP ingestion.⁶² Studies conducted in humans have demonstrated that NAC improves cerebral blood flow⁶³ and improves oxygen delivery and extraction in the hepatic-splanchnic circulation.^{64–66} A retrospective study analyzed 100 consecutive patients with acute liver failure from APAP poisoning who were treated with intravenous NAC in a 20-hour regimen or intravenous dextrose between 10 and 36 hours after APAP ingestion. Results were notable for lower mortality in the NAC-treated group (37%) versus the untreated controls (58%; $P < 0.05$). Progression to grade III or IV coma in the NAC group was also less frequent (51%) in comparison with

the untreated group (97%; $P < 0.05$).⁶⁷ In a prospective, randomized study, 50 consecutive patients with APAP-induced fulminant hepatic failure were randomized to receive either intravenous NAC treatment or intravenous dextrose. Patients were treated until recovery or death. Survival was significantly better in the NAC-treated group (48%) than in the untreated group (20%). Cerebral edema was also less likely in the NAC-treated group (40%), and there were fewer episodes of hypotension requiring inotropes in the NAC-treated group (48%) than among untreated controls (68% and 80%, respectively).⁶⁸ In light of these studies, most would recommend the administration of NAC, regardless of how late or how profound the presentation of APAP poisoning is.

The reported duration of NAC therapy is variable, with various lengths of administration in multiple studies and protocols ranging from 20 to 72 hours of treatment. An intravenous 20-hour NAC regimen is supported by extensive clinical experience.⁶⁹ However, no major difference has been found when NAC is initiated within 8 hours of APAP ingestion,⁷⁰ whereas longer courses of treatment have been found to be more effective in patients with a late presentation⁷⁰ and in patients who present with acute liver failure.^{31,67}

The specific route of administration does not have an impact on the effectiveness as NAC can be administered orally or intravenously. Although more adverse drug reactions have been reported with the intravenous route (up to 25%), these are more likely to be dose-related or rate-related rather than route-related.^{71,72} Intravenous administration is also recommended if oral ingestion or absorption is impaired or hazardous or if the patient already has altered mental status. A study analyzing 2540 patients with APAP poisoning from 1976 to 1985 reported an oral 72-hour NAC regimen to be more effective than the standard intravenous 20-hour regimen in reducing hepatotoxicity in patients who were treated late (beyond 10 hours of APAP ingestion).³¹ However, these findings may have been related to the larger doses of NAC used or the longer duration of treatment, rather than the difference of the route of administration of NAC per se. Intravenous NAC is administered as 150 mg/kg over 15 minutes, then 50 mg/kg over 4 hours, and then 100 mg/kg over 16 hours. Oral NAC is administered as 140 mg/kg as a loading dose and then 70 mg/kg every 4 hours for another 17 doses. Vomiting within an hour of oral NAC administration requires repeat dosing.

Oral Charcoal

Oral activated charcoal is also useful if administered within 4 hours of APAP ingestion. It may be used beyond the initial 4 hours in the presence of delayed gastric emptying or APAP absorption (eg, with coingestants that reduce gut motility). It has been found to adsorb APAP, resulting in reduced absorption of the drug. In patients with known or suspected APAP overdose who present within 4 hours of ingestion, administration of activated charcoal is recommended as first-line therapy, even before NAC.^{2,73} A study comparing the use of gastric lavage, syrup of ipecac, and oral activated charcoal among 20 patients found that activated charcoal produced a greater lowering of mean serum APAP levels than other interventions.⁷⁴ It has the best risk-benefit ratio in comparison with other decontaminants.⁷⁵ Oral activated charcoal is administered as a single oral dose of 1 g/kg. There is no benefit to the use of divided doses.

Cimetidine

The use of cimetidine to treat APAP toxicity was based on the observation that it is also metabolized by the cytochrome p450 2E1 pathway, which would theoretically lead to competitive inhibition of the enzyme and reduce APAP metabolism to NAPQI. In an early study, however, Slattery et al.⁷⁶ reported that the administration of 300 mg of cimetidine every 6 hours to 13 subjects after 8 hours of APAP ingestion did not alter APAP metabolism or APAP elimination or reduce alanine transferase or aspartate transferase levels. It was suggested that this lack of effect reflected the late administration of cimetidine after APAP ingestion.⁷⁷

Dialysis

The role of extracorporeal elimination in APAP intoxication is controversial, and the data are scanty.²² Hemodialysis has been used in severe APAP hepatotoxicity as the drug is dialyzable. However, results have not shown that hemodialysis prevents or reduces the risk of hepatotoxicity.²² Possible reasons for this include the high volume of distribution of APAP,⁷⁸ the fact that hepatotoxicity occurs early in overdose during the first pass through the liver,^{79,80} and the fact that toxic metabolites are intrahepatic and thus are not removed by hemodialysis.⁸¹ However, it may be useful in treating hyperammonemia associated with encephalopathy or in patients with persistently high plasma levels of APAP.⁸²

As with hemodialysis, although APAP elimination is enhanced by charcoal hemoperfusion, outcome benefit has not been demonstrated.^{75,83,84}

Notably, most reports have described hemodialysis or hemoperfusion after biochemical evidence of liver injury or after onset of acute liver failure.

Few studies have evaluated liver dialysis specifically for APAP hepatotoxicity. In an early case series, Detry et al.⁸⁵ reported that 8 patients with APAP poisoning who fulfilled the King's College criteria for poor prognosis were treated with bioartificial liver support every 6 hours, in addition to standard medical treatment with NAC and intracranial measurements. Three of the patients were bridged successfully to liver transplantation, whereas the remaining 5 recovered spontaneously. In another study, Ash et al.⁸⁶ described in 10 patients a possible benefit from liver dialysis with extracorporeal sorbent-based devices more than 10 hours after APAP ingestion. All 10 patients subsequently recovered, with liver enzymes starting to normalize after 24 hours of treatment, and they were discharged between 3 and 7 days after overdose. None of the patients required a liver transplant. However, none of these reported patients had fulfilled King's College criteria for poor prognostic outcomes.

Liver Transplant

All patients with APAP-induced hepatic failure should be referred to a transplant center for further management. Care revolves around aggressive supportive management and continued intravenous NAC administered until recovery or transplant. A Cochrane review of 10 studies comparing mortality among patients who had a liver transplant and similar patients who did not found that 19 of 67 patients (28%) with a transplant died versus 121 of 180 patients (67%) who were not transplanted among patients who had fulfilled the King's College criteria.⁷⁵ Data on outcomes after successful liver transplantation for APAP-induced acute liver failure have been scanty, and follow-up periods at the time of publication have been variable.^{43,87,88} Bernal et al.⁸⁸ described the outcomes of 548 patients with APAP-induced hepatotoxicity from their center over a 7-year period. Only 35% of those who had fulfilled transplant criteria (by the King's College criteria) underwent a transplant, of whom 75% survived to leave the hospital. Survival was greatest among those who received unreduced grafts and did not show signs of early graft dysfunction. Rapid and lethal clinical deterioration was the most frequent reason for not listing patients who had fulfilled clinical criteria for transplantation. Only in a small proportion of patients was an adverse psychiatric history implicated in failure to list the patient for liver transplantation. However, 12% (n = 4) of the transplanted patients

who had survived to leave the hospital died at a median of 8 months following transplantation from deliberate self-harm. Of these, 2 had taken further overdoses, 1 committed suicide by violent means, and 1 died of graft failure from deliberate noncompliance.

REGULATION OF ACETAMINOPHEN

Legislative efforts in the United Kingdom in 1998 have resulted in a more aggressive approach toward limiting overdosing by making over-the-counter APAP available only in limited quantities (typically 16 tablets or capsules) and requiring blister packaging (making impulsive consumption of large doses less likely). Larger quantities of APAP can be obtained only with a prescription. Comparing the prelegislation and postlegislation periods, a large referral hospital in London saw a 74% decline in APAP-related deaths or transplants,⁸⁹ and most other studies have also indicated a significant decline in hospital admissions, deaths, and liver transplants.⁹⁰⁻⁹²

In September 2002, the US Food and Drug Administration held a 2-day meeting of its non-prescription-drug advisory committee to discuss the issue of APAP hepatotoxicity. The outcome proposals were to improve the clarity of the package labels and to consider a patient and physician education program. However, limiting the amount of APAP that can be purchased (eg, by blister packs or with a prescription only) was not addressed. To date, APAP remains on the US market in bottles containing as many as 500 tablets, each containing 500 mg.⁵

SUMMARY

APAP overdose is a major cause of acute liver failure. Although early administration of the specific antidote NAC can prevent severe hepatotoxicity, there are still patients who may progress to acute liver failure, requiring a liver transplant.

DISCLOSURES

Potential conflict of interest: Paul Martin was previously a consultant for McNeil Pharmaceuticals.

REFERENCES

- Davidson DG, Eastham WN. Acute liver necrosis following overdose of paracetamol. *Br Med J* 1966; 2: 497-499.
- Polson J, Lee WM. AASLD position paper: the management of acute liver failure. *Hepatology* 2005; 41: 1179-1197.
- Watson WA, Litovitz TL, Klein-Schwartz W, et al. 2003 annual report of the American Association of Poison Control Centers Toxic Exposure Surveillance System. *Am J Emerg Med* 2004; 22: 335-404.
- Lee WM. Acetaminophen and the U.S. Acute Liver Failure Study Group: lowering the risks of hepatic failure. *Hepatology* 2004; 40: 6-9.
- Lee WM. Acetaminophen toxicity: changing perceptions on a social/medical issue. *Hepatology* 2007; 46: 966-970.
- Blei AT. Brain edema in acute liver failure. *Crit Care Clin* 2008; 24: 99-114, ix.
- Larson AM, Polson J, Fontana RJ, et al. Acetaminophen-induced acute liver failure: results of a United States multicenter, prospective study. *Hepatology* 2005; 42: 1364-1372.
- Davern TJ II, James LP, Hinson JA, et al. Measurement of serum acetaminophen-protein adducts in patients with acute liver failure. *Gastroenterology* 2006; 130: 687-694.
- Khandelwal N, James LP, Simpson PM, et al. Identifying acetaminophen toxicity amongst indeterminate causes of acute liver failure [abstract]. *Gastroenterology* 2008; 134: SA-752.
- Forrest JA, Clements JA, Prescott LF. Clinical pharmacokinetics of paracetamol. *Clin Pharmacokinet* 1982; 7: 93-107.
- Bizovi KE, Aks SE, Paloucek F, et al. Late increase in acetaminophen concentration after overdose of Tylenol Extended Relief. *Ann Emerg Med* 1996; 28: 549-551.
- Cetarik EW, Dart RC, Hurlbut KM, et al. Tylenol Extended Relief overdose. *Ann Emerg Med* 1997; 30: 104-108.
- Tighe TV, Walter FG. Delayed toxic acetaminophen level after initial four hour nontoxic level. *J Toxicol Clin Toxicol* 1994; 32: 431-434.
- Douglas DR, Sholar JB, Smilkstein MJ. A pharmacokinetic comparison of acetaminophen products (Tylenol Extended Relief vs regular Tylenol). *Acad Emerg Med* 1996; 3: 740-744.
- Benson GD, Koff RS, Tolman KG. The therapeutic use of acetaminophen in patients with liver disease. *Am J Ther* 2005; 12: 133-141.
- Lewis RK, Paloucek FP. Assessment and treatment of acetaminophen overdose. *Clin Pharm* 1991; 10: 765-774.
- Prescott LF. Paracetamol overdosage. Pharmacological considerations and clinical management. *Drugs* 1983; 25: 290-314.
- Makin AJ, Wendon J, Williams R. A 7-year experience of severe acetaminophen-induced hepatotoxicity (1987-1993). *Gastroenterology* 1995; 109: 1907-1916.
- Corcoran GB, Mitchell JR, Vaishnav YN, Horning EC. Evidence that acetaminophen and N-hydroxyacetaminophen form a common arylating intermediate, N-acetyl-p-benzoquinoneimine. *Mol Pharmacol* 1980; 18: 536-542.
- Bessemers JG, Vermeulen NP. Paracetamol (acetaminophen)-induced toxicity: molecular and biochemical mechanisms, analogues and protective approaches. *Crit Rev Toxicol* 2001; 31: 55-138.

21. Smilkstein MJ. Acetaminophen. In: Goldfrank LR, Flomenbaum NE, Lewin NA, eds. *Goldfrank's Toxicologic Emergencies*. Stamford, CT: Appleton & Lange 1998; 541.
22. McBride PV, Rumack BH. Acetaminophen intoxication. *Semin Dial* 1992; 5: 292–297.
23. Suzuki A, Yuen NA, Papay JJ, et al. An Aers database evaluation of influence of co-medications on acetaminophen-associated liver injury. *Gastroenterology* 2008; 134: SA–758.
24. Polson J, Orsulik PJ, Wiens F, et al. Elevated bilirubin may cause false positive acetaminophen levels in hepatitis patients [abstract]. *Hepatology* 2004; 40: 496A.
25. Bertholf RL, Johannsen LM, Bazooband A, Mansouri V. False-positive acetaminophen results in a hyperbilirubinemic patient. *Clin Chem* 2003; 49: 695–698.
26. Beuhler MC, Curry SC. False positive acetaminophen levels associated with hyperbilirubinemia. *Clin Toxicol (Phila)* 2005; 43: 167–170.
27. Read RB, Tredger JM, Williams R. Analysis of factors responsible for continuing mortality after paracetamol overdose. *Hum Toxicol* 1986; 5: 201–206.
28. Ambre J, Alexander M. Liver toxicity after acetaminophen ingestion. Inadequacy of the dose estimate as an index of risk. *JAMA* 1977; 238: 500–501.
29. Rumack BH, Matthew H. Acetaminophen poisoning and toxicity. *Pediatrics* 1975; 55: 871–876.
30. Rumack BH. Acetaminophen hepatotoxicity: the first 35 years. *J Toxicol Clin Toxicol* 2002; 40: 3–20.
31. Smilkstein MJ, Knapp GL, Kulig KW, Rumack BH. Efficacy of oral N-acetylcysteine in the treatment of acetaminophen overdose. Analysis of the national multicenter study (1976 to 1985). *N Engl J Med* 1988; 319: 1557–1562.
32. Wootton FT, Lee WM. Acetaminophen hepatotoxicity in the alcoholic. *South Med J* 1990; 83(10):1047–1049.
33. Whitcomb DC, Block GD. Association of acetaminophen hepatotoxicity with fasting and ethanol use. *JAMA* 1994; 272: 1845–1850.
34. Rumack BH, Holtzman J, Chase HP. Hepatic drug metabolism and protein malnutrition. *J Pharmacol Exp Ther* 1973; 186: 441–446.
35. Dart RC, Kuffner EK, Rumack BH. Treatment of pain or fever with paracetamol (acetaminophen) in the alcoholic patient: a systematic review. *Am J Ther* 2000; 7: 123–134.
36. Prescott LF. Paracetamol, alcohol and the liver. *Br J Clin Pharmacol* 2000; 49: 291–301.
37. Kuffner EK, Dart RC, Bogdan GM, et al. Effect of maximal daily doses of acetaminophen on the liver of alcoholic patients: a randomized, double-blind, placebo-controlled trial. *Arch Intern Med* 2001; 161(10):2247–2252.
38. Prescott LF. Therapeutic misadventure with paracetamol: fact or fiction? *Am J Ther* 2000; 70:991–14.
39. Dart RC, Bailey E. Does therapeutic use of acetaminophen cause acute liver failure? *Pharmacotherapy* 2007; 27: 1219–1230.
40. Shayiq RM, Roberts DW, Rothstein K, et al. Repeat exposure to incremental doses of acetaminophen provides protection against acetaminophen-induced lethality in mice: an explanation for high acetaminophen dosage in humans without hepatic injury. *Hepatology* 1999; 29: 451–463.
41. Schiodt FV, Rochling FA, Casey DL, Lee WM. Acetaminophen toxicity in an urban county hospital. *N Engl J Med* 1997; 337: 1112–1117.
42. Lee WM. Acute liver failure in the United States. *Semin Liver Dis* 2003; 23: 217–226.
43. Mutimer DJ, Ayres RC, Neuberger JM, et al. Serious paracetamol poisoning and the results of liver transplantation. *Gut* 1994; 35: 809–814.
44. Anand AC, Nightingale P, Neuberger JM. Early indicators of prognosis in fulminant hepatic failure: an assessment of the King's criteria. *J Hepatol* 1997; 26: 62–68.
45. O'Grady JG, Alexander GJ, Hayllar KM, Williams R. Early indicators of prognosis in fulminant hepatic failure. *Gastroenterology* 1989; 97: 439–445.
46. Bernal W, Donaldson N, Wyncoll D, Wendon J. Blood lactate as an early predictor of outcome in paracetamol-induced acute liver failure: a cohort study. *Lancet* 2002; 359: 558–563.
47. Pereira LM, Langley PG, Hayllar KM, et al. Coagulation factor V and VIII/V ratio as predictors of outcome in paracetamol induced fulminant hepatic failure: relation to other prognostic indicators. *Gut* 1992; 33: 98–102.
48. Mitchell I, Bihari D, Chang R, et al. Earlier identification of patients at risk from acetaminophen-induced acute liver failure. *Crit Care Med* 1998; 26: 279–284.
49. Schmidt LE, Dalhoff K. Serum phosphate is an early predictor of outcome in severe acetaminophen-induced hepatotoxicity. *Hepatology* 2002; 36: 659–665.
50. Chung PY, Sitrin MD, Te HS. Serum phosphorus levels predict clinical outcome in fulminant hepatic failure. *Liver Transpl* 2003; 9: 248–253.
51. Ng KL, Davidson JS, Bathgate AJ. Serum phosphate is not a reliable early predictor of outcome in paracetamol induced hepatotoxicity. *Liver Transpl* 2004; 10: 158–159.
52. Gregory B, Odina CV, Moe O, et al. Abnormal urinary excretion of phosphate is responsible for hypophosphataemia in acute liver failure [abstract]. *Gastroenterology* 2008; 134: SA–800.
53. Schmidt LE, Larsen FS. MELD score as a predictor of liver failure and death in patients with acetaminophen-induced liver injury. *Hepatology* 2007; 45: 789–796.
54. Schmidt LE, Dalhoff K. Hyperamylasaemia and acute pancreatitis in paracetamol poisoning. *Aliment Pharmacol Ther* 2004; 20: 173–179.
55. Schiodt FV, Bondesen S, Tygstrup N, Christensen E. Prediction of hepatic encephalopathy in paracetamol overdose: a prospective and validated study. *Scand J Gastroenterol* 1999; 34: 723–728.
56. Harrison PM, O'Grady JG, Keays RT, et al. Serial prothrombin time as prognostic indicator in paracetamol induced fulminant hepatic failure. *BMJ* 1990; 301: 964–966.
57. Lin JH, Levy G. Sulfate depletion after acetaminophen administration and replenishment by infusion of sodium sulfate or N-acetylcysteine in rats. *Biochem Pharmacol* 1981; 30: 2723–2725.
58. Prescott LF, Illingworth RN, Critchley JA, et al. Intravenous N-acetylcysteine: the treatment of choice for paracetamol poisoning. *Br Med J* 1979; 2: 1097–1100.
59. Smilkstein MJ, Bronstein AC, Linden C, et al. Acetaminophen overdose: a 48-hour intravenous N-acetylcysteine treatment protocol. *Ann Emerg Med* 1991; 20: 1058–1063.
60. Cuzzocrea S, Costantino G, Mazzon E, Caputi AP. Protective effect of N-acetylcysteine on multiple organ failure induced by zymosan in the rat. *Crit Care Med* 1999; 27: 1524–1532.

61. Bajt ML, Knight TR, Lemasters JJ, Jaeschke H. Acetaminophen-induced oxidant stress and cell injury in cultured mouse hepatocytes: protection by N-acetylcysteine. *Toxicol Sci* 2004; 80: 343–349.
62. Rowden AK, Norvell J, Eldridge DL, Kirk MA. Acetaminophen poisoning. *Clin Lab Med* 2006; 26: 49–65, viii.
63. Wendon JA, Harrison PM, Keays R, Williams R. Cerebral blood flow and metabolism in fulminant liver failure. *Hepatology* 1994; 19: 1407–1413.
64. Spies CD, Reinhart K, Witt I, et al. Influence of N-acetylcysteine on indirect indicators of tissue oxygenation in septic shock patients: results from a prospective, randomized, double-blind study. *Crit Care Med* 1994; 22: 1738–1746.
65. Harrison PM, Wendon JA, Gimson AE, et al. Improvement by acetylcysteine of hemodynamics and oxygen transport in fulminant hepatic failure. *N Engl J Med* 1991; 324: 1852–1857.
66. Devlin J, Ellis AE, McPeake J, et al. N-acetylcysteine improves indocyanine green extraction and oxygen transport during hepatic dysfunction. *Crit Care Med* 1997; 25: 236–242.
67. Harrison PM, Keays R, Bray GP, et al. Improved outcome of paracetamol-induced fulminant hepatic failure by late administration of acetylcysteine. *Lancet* 1990; 335: 1572–1573.
68. Keays R, Harrison PM, Wendon JA, et al. R. Intravenous acetylcysteine in paracetamol induced fulminant hepatic failure: a prospective controlled trial. *BMJ* 1991; 303: 1026–1029.
69. Prescott LF, Illingworth RN, Critchley JA, Proudfoot AT. Intravenous N-acetylcysteine: still the treatment of choice for paracetamol poisoning. *Br Med J* 1980; 280: 46–47.
70. Buckley NA, Whyte IM, O'Connell DL, Dawson AH. Oral or intravenous N-acetylcysteine: which is the treatment of choice for acetaminophen (paracetamol) poisoning? *J Toxicol Clin Toxicol* 1999; 37: 759–767.
71. Yip L, Dart RC, Hurlbut KM. Intravenous administration of oral N-acetylcysteine. *Crit Care Med* 1998; 26: 40–43.
72. Bailey B, Amre DK, Gaudreault P. Fulminant hepatic failure secondary to acetaminophen poisoning: a systematic review and meta-analysis of prognostic criteria determining the need for liver transplantation. *Crit Care Med* 2003; 31: 299–305.
73. Green R, Grierson R, Sitar DS, Tenenbein M. How long after drug ingestion is activated charcoal still effective? *J Toxicol Clin Toxicol* 2001; 39: 601–605.
74. Underhill TJ, Greene MK, Dove AF. A comparison of the efficacy of gastric lavage, ipecacuanha and activated charcoal in the emergency management of paracetamol overdose. *Arch Emerg Med* 1990; 7: 148–154.
75. Brok J, Buckley N, Gluud C. Interventions for paracetamol (acetaminophen) overdose. *Cochrane Database Syst Rev* 2006; CD003328.
76. Slattery JT, McRorie TI, Reynolds R, et al. Lack of effect of cimetidine on acetaminophen disposition in humans. *Clin Pharmacol Ther* 1989; 46: 591–597.
77. Burkhart KK, Janco N, Kulig KW, Rumack BH. Cimetidine as adjunctive treatment for acetaminophen overdose. *Hum Exp Toxicol* 1995; 14: 299–304.
78. Marbury TC, Wang LH, Lee CS. Hemodialysis of acetaminophen in uremic patients. *Int J Artif Organs* 1980; 3: 263–266.
79. Matthew H. Acute acetaminophen poisoning. *Clin Toxicol* 1973; 6: 9–11.
80. Prescott LF. Haemodialysis in paracetamol self-poisoning. *Lancet* 1972; 2: 652.
81. N-acetylcysteine in the treatment of acetaminophen overdose. *N Engl J Med* 1989; 320: 1417–1418.
82. Williams R. Hepatic encephalopathy. *JR Coll Physicians Lond* 1973; 8: 63–74.
83. Gazzard BG, Willson RA, Weston MJ, et al. Charcoal haemoperfusion for paracetamol overdose. *Br J Clin Pharmacol* 1974; 1: 271.
84. O'Grady JG, Gimson AE, O'Brien CJ, et al. Controlled trials of charcoal hemoperfusion and prognostic factors in fulminant hepatic failure. *Gastroenterology* 1988; 94: 1186–1192.
85. Detry O, Arkadopoulos N, Ting P, et al. Clinical use of a bioartificial liver in the treatment of acetaminophen-induced fulminant hepatic failure. *Am Surg* 1999; 65: 934–938.
86. Ash SR, Caldwell CA, Singer GG, et al. Treatment of acetaminophen-induced hepatitis and fulminant hepatic failure with extracorporeal sorbent-based devices. *Adv Ren Replace Ther* 2002; 9: 42–53.
87. O'Grady JG, Wendon J, Tan KC, et al. Liver transplantation after paracetamol overdose. *BMJ* 1991; 303: 221–223.
88. Bernal W, Wendon J, Rela M, et al. Use and outcome of liver transplantation in acetaminophen-induced acute liver failure. *Hepatology* 1998; 27: 1050–1055.
89. Sheen CL, Dillon JF, Bateman DN, et al. Paracetamol pack size restriction: the impact on paracetamol poisoning and the over-the-counter supply of paracetamol, aspirin and ibuprofen. *Pharmacoepidemiol Drug Saf* 2002; 11: 329–331.
90. Hawton K, Townsend E, Deeks J, et al. Effects of legislation restricting pack sizes of paracetamol and salicylate on self poisoning in the United Kingdom: before and after study. *BMJ* 2001; 322: 1203–1207.
91. Hawton K, Simkin S, Deeks J, et al. UK legislation on analgesic packs: before and after study of long term effect on poisonings. *BMJ* 2004; 329: 1076.
92. Wilkinson S, Taylor G, Templeton L, et al. Admissions to hospital for deliberate self-harm in England 1995–2000: an analysis of hospital episode statistics. *J Public Health Med* 2002; 24: 179–183.

Healthcare-Associated Infections: Epidemiology, Prevention, and Therapy

Rupali K. Doshi, Gopi Patel, Richard MacKay, and Frances Wallach

Mount Sinai School of Medicine, New York, NY

ABSTRACT

Reducing nosocomial infection rates is a major component of healthcare improvement. This article reviews the epidemiology, prevention, and therapy for some of the most common healthcare-associated infections, including central line-associated bloodstream infections and catheter-associated urinary tract infections, and 3 common organisms: methicillin-resistant *Staphylococcus aureus*, multidrug-resistant gram-negative bacteria, and *Clostridium difficile*. **Mt Sinai J Med 76:84–94, 2009.** © 2009 Mount Sinai School of Medicine

Key Words: *Acinetobacter baumannii*, bloodstream infection, *Clostridium difficile*, extended-spectrum β -lactamase, healthcare-associated infection, methicillin-resistant *Staphylococcus aureus*, multidrug-resistant gram-negative bacteria, nosocomial, quality improvement, urinary tract infection.

Healthcare-associated infections are a significant source of patient morbidity and mortality. Individual providers and health systems should be actively involved in the surveillance, diagnosis, and early management of healthcare-associated infections with the goal of reducing the risk of preventable complications during the hospital stay. Nosocomial infections are also associated with increased healthcare expenditures, and as of 2008, the Centers for Medicare and Medicaid Services has chosen to withhold reimbursement for certain hospital-acquired conditions, including catheter-associated urinary tract infections (CAUTIs) and central line-associated bloodstream infections (CLABSI).¹ Because of these clinical, ethical, and financial factors, healthcare providers,

regulatory agencies, and consumer advocacy groups are paying increased attention to the prevention of healthcare-associated infections in the acute care setting.²

Hospitals across the United States submit surveillance data to the National Healthcare Safety Network, and the Centers for Disease Control and Prevention (CDC) performs analysis on the aggregated data.³ Although national surveillance is ongoing, individual practitioner knowledge of institution-specific surveillance data, recognition of common clinical syndromes, and effective preventive measures are essential for the prevention, early detection, and treatment of healthcare-associated infections.⁴

This article is focused on the epidemiology and prevention of some of the most common infectious complications seen in hospitals, including CLABSI and CAUTI. Additionally, the epidemiology, prevention, and therapy for some of the most prevalent healthcare-associated pathogens, including methicillin-resistant *Staphylococcus aureus* (MRSA), multidrug-resistant gram-negative bacteria, and *Clostridium difficile*, are discussed.

CENTRAL LINE-ASSOCIATED BLOODSTREAM INFECTIONS

Approximately 200,000 CLABSI occur annually in US hospitals, and 500 to 4000 US deaths occur annually because of CLABSI.^{5,6} CLABSI have been associated with an in-hospital mortality rate of 12% to 35%, prolonged hospitalizations, and complications associated with the treatment of these infections.^{7–9} The National Healthcare Safety Network collects surveillance data on CLABSI in intensive care units (ICUs), certain specialty care areas (eg, transplant units and inpatient dialysis units), and other surgical and medical wards. A central line provides intravenous access that terminates in one of the large veins. A CLABSI, according to CDC surveillance criteria, is a primary bloodstream infection with a

Address Correspondence to:

Rupali K. Doshi

Mount Sinai School of Medicine,
New York, NY

Email: rupali.k.doshi@gmail.com

central line in place at the time of infection or up to 48 hours previously. A primary bloodstream infection indicates that the infection is not attributable to infection at another site.¹⁰ A CLABSI must be a laboratory-confirmed primary bloodstream infection, meeting at least 1 of 2 criteria: 1 or more positive blood cultures with a recognized pathogen and/or 2 or more positive blood cultures for skin contaminants (eg, coagulase-negative *Staphylococcus*) with fever, chills, or hypotension (clinical sepsis).⁵

Over half of all CLABSIs are preventable with the implementation of evidence-based infection-reducing practices for the insertion and maintenance of central lines.¹¹ During placement of the catheter, several procedural methods have been associated with a reduction in CLABSI rates. These include disinfecting the prospective catheter site with a 2% chlorhexidine-based antiseptic solution (preferred over povidone-iodine); using maximal barrier precautions, including sterile gowns, sterile gloves, masks, caps, and large sterile drapes; employing appropriate hand hygiene; and educating healthcare providers who insert and care for central venous catheters.¹² The Institute for Healthcare Improvement promotes the use of the central line bundle, which consists of 5 key components: hand hygiene, maximal barrier precautions, chlorhexidine skin antisepsis, optimal catheter site selection (a subclavian vein is preferred for nontunneled catheters if there are no contraindications), and daily review of central line necessity with removal of unnecessary lines.¹³

For maintenance of central venous catheters, regular cleansing of the catheter site by trained staff, avoidance of antibiotic ointments, and daily washing of the patient's body with a 2% chlorhexidine solution can be used.^{12,14} A chlorhexidine-impregnated sponge placed at the catheter's site of entry into the skin may prevent some CLABSIs.¹⁵ A systematic review concluded that the use of antimicrobial-impregnated central lines should be limited to settings with high rates of CLABSIs.¹⁶ Cuffed and tunneled catheters have lower rates of CLABSIs than nontunneled catheters, and lower infection rates have been associated with the use of polytetrafluoroethylene-coated and polyurethane-coated catheters, although they are more expensive.⁹ Peripherally inserted central catheters may have a slightly lower risk of CLABSIs in comparison with central lines placed in the neck or chest because the arm is less heavily colonized with skin bacteria than those areas; however, peripherally inserted central catheter lines are associated with higher rates of bloodstream infections in inpatients compared with outpatients.⁹ There is little evidence that routine replacement of central lines reduces infection rates.¹⁷

Common pathogens associated with CLABSIs include coagulase-negative *Staphylococcus*, *S. aureus*, *Enterococcus* species, *Candida* species, and gram-negative bacilli.¹⁷

CATHETER-ASSOCIATED URINARY TRACT INFECTIONS

Urinary tract infections compose 40% of all healthcare-associated infections, and the majority of these are related to the use of urinary catheters.¹⁸ CAUTIs generally occur because of contamination during catheter insertion or the presence of a biofilm on the surface of the indwelling catheter. A biofilm is a population of bacteria in an extracellular matrix that develops on the surfaces of devices used in humans.¹⁹

The CDC definition of a CAUTI is a urinary tract infection that occurs with a urinary catheter in place or within 7 days of its placement; the definition includes both clinical and laboratory criteria and can be met one of two ways (Table 1).⁵ Some have criticized this definition because hospitalized patients may not demonstrate some of the requisite physical signs and symptoms.¹⁸ Also, the inclusion of a physician diagnosis of a urinary tract infection is subjective and may be dependent on individual practice patterns and opinion. This may erroneously alter true CAUTI rates.

In order to reduce the risk of CAUTIs, the urinary catheter should be inserted with a sterile technique, a closed system for collection should be in place, and the collection bag should be drained by gravity. There are several system-based strategies that may reduce the incidence of CAUTIs. Antimicrobial urinary catheters (eg, nitrofurazone or silver) have been shown to reduce the incidence of CAUTIs, but because of the increased costs, some hospitals have determined that they should be used only in high-risk populations, such as ICU or transplant patients.²⁰ Computer-based reminders for catheter removal or interdisciplinary rounds on patients with catheters to determine their ongoing necessity can also be effective. Other evidence-based interventions include bladder ultrasound to determine the postvoid residual and the substitution of condom catheters or suprapubic catheters for Foley catheters in men.^{21,22} Management strategies without evidence include the use of antimicrobial-impregnated urine collection bags, frequent cleaning of the meatus, and bladder irrigation.²⁰

Saint et al.²⁰ found that roughly half of sampled hospitals had no system of measurement of catheter

Table 1. Centers for Disease Control and Prevention Definition for Catheter-Associated Urinary Tract Infections.

Clinical Criteria	Laboratory Data
1. At least 1 of the following symptoms: fever, urgency, frequency, dysuria, or suprapubic pain	Positive urine culture (> 100,000 colonies/mL with no more than 2 organisms)
2. At least 2 of the following symptoms: fever (> 38°C), urgency, frequency, dysuria, or suprapubic tenderness	At least 1 of the following: 1. Positive dipstick for leukocyte esterase and/or nitrate 2. Pyuria (urine specimen with ≥ 10 white blood cells/mm³ or ≥ 3 white blood cells/high-power field of unspun urine) 3. Organisms seen on Gram stain of unspun urine 4. At least 2 urine cultures with repeated isolation of the same uropathogen (gram-negative bacteria or <i>Staphylococcus saprophyticus</i>) with $\geq 10^2$ colonies/mL in nonvoided specimens 5. $\leq 10^5$ colonies/mL of a single uropathogen (gram-negative bacteria or <i>S. saprophyticus</i>) in a patient being treated with an effective antimicrobial agent for a urinary tract infection. 6. Physician diagnosis of a urinary tract infection. 7. Appropriate therapy for a urinary tract infection instituted by a physician

NOTE: The information for this table was taken from the *National Healthcare Safety Network Manual*.⁵

days or systems in place for catheter removal. In fact, bladder ultrasound, computer-based reminders, and antimicrobial-impregnated urinary catheters were used in less than one-third of the sampled hospitals.

Gram-negative bacteria remain the most common cause of CAUTIs. These bacteria are becoming increasingly resistant to commonly used oral antibiotics, and this makes the treatment of these infections progressively more complicated. Further discussion about gram-negative bacteria is included later in this article.

METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS*

MRSA is a major cause of healthcare-associated infections and is now endemic in most US hospitals. Over the past decade, the striking increase in the prevalence of healthcare-associated methicillin-resistant *S. aureus* (HA-MRSA) infections, the epidemic emergence of community-associated methicillin-resistant *S. aureus* (CA-MRSA) infections in otherwise healthy adults and children, and the failure of vancomycin to treat some laboratory-defined susceptible MRSA strains have heightened concerns and led to calls for more aggressive measures to control this pathogen.^{22–26}

Humans are the principal reservoir of *S. aureus*, usually through colonization of the anterior nares. Nasal colonization usually precedes infection.

Nosocomial spread occurs primarily through exposure to the hands of healthcare workers after they have been transiently colonized with *S. aureus* from contact with an infected or colonized patient. Contact with contaminated equipment or environmental surfaces has also been implicated in spread. Data from the 2001–2002 National Health and Nutrition Examination Survey estimated national prevalence rates of *S. aureus* and MRSA colonization of 32.4% and 0.8%, respectively.^{27,28}

Methicillin resistance is conferred by an extra-chromosomal genetic element, *mecA*, that encodes for penicillin binding protein 2a, an altered membrane protein with a reduced affinity for β -lactams.²⁸ HA-MRSA strains characteristically carry *mecA* subtypes I–III; they are resistant to all currently available β -lactams and have variable resistance patterns to other antimicrobials effective against gram-positive cocci. The *mecA* gene is part of a larger mobile gene cassette, the staphylococcal chromosomal cassette (SCC), and it has been postulated that the clonal expansion of strains with this element drives MRSA spread.²⁹ Once *S. aureus* is isolated in the clinical laboratory, methicillin resistance is determined with oxacillin or cefoxitin disk-diffusion techniques. Chromogenic agar incubation and newer molecular methods that detect the *mecA* gene by real-time polymerase chain reaction shorten the time to diagnosis of MRSA.^{30,31}

HA-MRSA infections have steadily increased since the first reported nosocomial outbreak in 1978 and are now endemic in most US hospitals.³²

Between 1999 and 2005, hospitalizations for serious MRSA infections more than doubled, and by 2003, 64% of *S. aureus* isolates from ICUs were methicillin-resistant.^{33,34} Using CDC national surveillance data, Klevens et al.²² estimated that in 2005 there were 91,360 invasive MRSA infections and 18,650 MRSA-associated deaths in the United States. Cases were defined by the isolation of MRSA from a normally sterile site and further characterized as bacteremia (75.2%), pneumonia (13.3%), cellulitis (9.7%), osteomyelitis (7.5%), endocarditis (6.3%), and septic shock (4.3%). Crude mortality rates were highest in those with septic shock and pneumonia. Eighty-five percent were identified as healthcare-associated infections, as defined by established risk factors for MRSA infection. Of these HA-MRSA infections, nearly 70% had their onset in the community. Epidemiological studies have identified hospitalization in an ICU, age >65 years, length of hospital stay, exposure to colonized or infected patients, central venous catheters, mechanical ventilation, and antibiotic therapy as important risk factors associated with MRSA infection.⁴⁵⁻⁴⁶

CA-MRSA infections, defined by absence of the aforementioned traditional healthcare-associated risk factors, have increased dramatically in the United States over the past 10 years. Although initially reported as a fatal pathogen in 4 upper Midwest pediatric patients, CA-MRSA is now most closely associated with severe skin and soft tissue infections, including necrotizing fasciitis.^{25,37}

Genotypically, CA-MRSA isolates are identified by unique pulsed-field gel electrophoresis patterns. The most frequently isolated CA-MRSA clone is USA300, which carries a novel smaller SCC, *mecA* type IV, and often expresses Panton-Valentine leukocidin.³⁸ The exact role of Panton-Valentine leukocidin in the increased virulence of these isolates has not yet been elucidated.³⁹ Unlike hospital-associated clones (HA-MRSA), CA-MRSA isolates often retain susceptibility to clindamycin, tetracycline, rifampin, and trimethoprim-sulfamethoxazole. The smaller SCC, *mec* IV, contains fewer resistance genes than its larger counterparts in HA-MRSA (types I-III). Phenotypically, these strains are linked primarily to severe skin and soft tissue infections, and initial studies defined at-risk groups in populations with close person-to-person contact, including athletic teams, incarcerated populations, daycare centers, and men who have sex with men.^{37,40} Infections are now commonly seen in otherwise healthy individuals outside these risk groups. A survey of 11 US university-hospital emergency departments in 2004 found an overall prevalence of MRSA among skin and

soft tissue infections of 59% (range, 15–74%). Ninety-seven percent of isolates were genotype USA300 and contained SCC *mecA* type IV.⁴¹ One participating hospital reported that this CA-MRSA genotype was isolated in 28% of HA-MRSA bloodstream infections and 20% of nosocomial bloodstream infections (defined by a positive culture obtained >48 hours after admission).²⁴ This remarkable phenomenon of a multidrug-resistant organism introduced into the healthcare system from the community has now blurred the distinction between CA-MRSA and HA-MRSA.

Vancomycin remains the drug of choice for most serious, susceptible MRSA infections.⁴² The addition of an aminoglycoside shortens bacteremia but has no impact on outcome. Combination therapy with vancomycin plus an aminoglycoside and rifampin is indicated for prosthetic valve endocarditis.⁴³ The duration of therapy for serious, invasive disease is 4 weeks or longer. Rarely, an immunocompetent patient with a catheter-associated blood stream infection, early removal of the infected device, prompt clearance of bacteremia, no evidence of metastatic infection, and normal transthoracic echocardiogram can be considered for a shorter 2-week course of therapy. Whenever possible, potentially infected catheters and other foreign devices should be removed.²⁸ Dosing should provide a trough concentration of 10 to 15 µg/mL. Infectious disease consultation is recommended for treating *S. aureus* bacteremias because of the associated morbidity and high treatment failure rates.

Vancomycin treatment failures in susceptible MRSA strains are increasingly common, especially when the median inhibitory concentration exceeds 1 µg/mL. This discordance between in vitro susceptibility and clinical success led the Clinical Laboratory Standards Institute to lower the breakpoint for vancomycin susceptibility to ≤2 µg/mL.⁴⁴ Aggressive vancomycin dosing to reach trough concentrations of >15 µg/mL has not resulted in improved outcomes and may be associated with increased nephrotoxicity.^{45,46} The failure of vancomycin to eliminate susceptible strains of MRSA is felt to be linked to the emergence of resistant daughter bacteria termed *heteroresistant vancomycin S. aureus*. Currently, no reliable validated laboratory method exists to identify these heteroresistant vancomycin *S. aureus* isolates.⁴⁷

Alternatives to vancomycin include daptomycin, linezolid, and tigecycline, but published clinical trials involving serious MRSA infections are limited. One industry-sponsored study found daptomycin to be noninferior to standard therapy with an antistaphylococcal penicillin or vancomycin for

S. aureus bacteremia and right-sided endocarditis.⁴⁸ In 6 patients, daptomycin resistance emerged while they were on therapy. Degradation of daptomycin by a surfactant limits its effectiveness in MRSA pneumonia. In a retrospective, subgroup analysis of 2 studies comparing linezolid to vancomycin for the treatment of nosocomial pneumonia, survival rates were significantly higher in patients infected with MRSA who were treated with linezolid.⁴⁹ Prospective studies are needed to confirm those results. Prolonged therapy with linezolid can lead to reversible thrombocytopenia, and the drug has been linked to serotonin syndrome, especially when combined with antidepressants such as selective serotonin reuptake inhibitors. As the severity and complexity of invasive MRSA infections increase, clinicians will need to collaborate closely with clinical laboratory personnel and infectious disease specialists to optimize outcomes.

In the United States, guidelines for the control and prevention of hospital-associated MRSA infections include hand hygiene, contact isolation of colonized or infected patients, environmental disinfection, antibiotic stewardship, and active surveillance cultures of high-risk patients.⁵⁰ A recently published report showed a 70% decrease in hospital-associated MRSA disease prevalence coincident with a program of universal admission cultures to detect nasal carriage of MRSA and decolonization of affected patients.³¹ Huang and colleagues⁵¹ demonstrated a significant hospital-wide reduction in MRSA bacteremia after instituting active surveillance cultures limited to ICU patients. Universal implementation of surveillance remains controversial.⁵²

HEALTHCARE-ASSOCIATED GRAM-NEGATIVE INFECTIONS

Gram-negative bacilli are responsible for the majority of urinary tract infections and pneumonias in ICUs and close to a quarter of bloodstream infections in ICUs.⁵³ Although the proportion of bloodstream infections attributable to gram-negative bacteria has decreased, there is a trend toward increasing prevalence of antibiotic resistance. Common multidrug-resistant gram-negative bacilli associated with healthcare-associated infections include extended-spectrum β -lactamase (ESBL)-producing Enterobacteriaceae and *Acinetobacter* species.

Third-generation cephalosporins were developed in order to overcome antibiotic resistance conferred by β -lactamases, bacteria-derived enzymes capable of hydrolyzing earlier β -lactam antibiotics.

Escalating use of cephalosporins (eg, ceftriaxone, cefotaxime, and ceftazidime) has led to increasing clinical isolation of Enterobacteriaceae, especially *Klebsiella pneumoniae* and *Escherichia coli*, producing ESBLs.⁵⁴ These enzymes are capable of inactivating all cephalosporins, including cefepime, as well as aztreonam, thus limiting the antibiotics available to treat serious gram-negative infections.⁵⁵ Surveillance studies have demonstrated increasing clinical isolation of *K. pneumoniae* strains resistant to third-generation cephalosporins.^{56,57}

Various factors have been linked to the clinical isolation of ESBL-producing bacteria, including length of stay, ICU exposure, arterial and central venous catheter use, urinary catheter use, emergent abdominal surgeries, use of feeding tubes, hemodialysis, ventilator dependence, and previous exposure to antibiotics with gram-negative activity.^{58,59}

ESBL-producing bacteria pose a therapeutic challenge because of their complicated resistance profiles. The transferable plasmids encoding these enzymes often carry resistance genes for aminoglycosides and sulfonamides as well.⁵⁸ Susceptibility to fluoroquinolones is variable, and fluoroquinolone resistance in the setting of ESBL production appears to be increasing.^{60,61} Most ESBL-producing Enterobacteriaceae demonstrate susceptibility to carbapenems (ie, imipenem, meropenem, and ertapenem), and this makes these agents the antibiotics of choice in the setting of severe sepsis or known ESBL-producing gram-negative infections.⁶² Clinical experience also supports the use of carbapenems as first-line agents in these settings.⁶³ In vitro susceptibility profiles for ESBL-producing Enterobacteriaceae can be misleading and may suggest that a specific strain is susceptible to a given cephalosporin. Clinical failures have been associated with treating these infections with cephalosporins despite contrary susceptibility profiles.^{64–66} Failure to initiate appropriate antibiotic therapy (ie, carbapenems) has been linked to mortality in the setting of ESBL-producing *K. pneumoniae* and *E. coli* infections.^{67,68}

Increasing reliance on carbapenems has led to the emergence of carbapenem resistance in other gram-negative bacilli, notably *Pseudomonas aeruginosa*⁶⁹ and *Acinetobacter* species.^{70,71} Carbapenem resistance among Enterobacteriaceae (eg, *K. pneumoniae*, *Enterobacter* species, and *E. coli*) is relatively rare, but clinical isolation is increasing in frequency, especially in New York City.^{72–74} Infectious disease consultation should be enlisted to assist in the treatment of infections with these organisms because successful treatment has been anecdotal at best.

Acinetobacter has the ability to withstand harsh environmental conditions and to acquire multiple

antimicrobial resistance determinants. It has been associated with several outbreaks in ICUs and other healthcare settings.⁷⁵ Healthcare-associated pneumonia, including ventilator-associated pneumonia, is the most common type of infection attributed to these gram-negative bacilli.⁷⁶ Risk factors for *Acinetobacter* infection are similar to those cited for infections with ESBL-producing Enterobacteriaceae.⁷⁷ According to national surveillance data, the proportion of ICU-associated pneumonias due to *Acinetobacter* increased from 4% in 1986 to 7% in 2003.⁵⁷ *Acinetobacter* species have also been associated with bloodstream infections and meningitis, complicating neurosurgical procedures, especially in institutions with a higher prevalence of these bacteria.

Healthcare-associated isolates of *Acinetobacter baumannii* often demonstrate resistance to penicillins, cephalosporins, aminoglycosides, and fluoroquinolones. Resistance mechanisms vary and include β -lactamases, cephalosporinases, and alterations in outer membrane proteins as well as aminoglycoside-modifying enzymes and efflux pumps.⁷⁸ In institutions in which carbapenem resistance rates are low, monotherapy with either imipenem or meropenem is usually effective in treating serious *Acinetobacter* infections. If there are supportive susceptibility data, ceftazidime and cefepime can be used in the treatment of less serious *A. baumannii* infections. Similarly to other multidrug-resistant organisms, prior antibiotic exposure has been linked to increasing antibiotic resistance.^{71,79,80}

Previous carbapenem use and high institutional rates of carbapenem utilization have been associated with the recovery of carbapenem-resistant strains of *Acinetobacter* species.^{79–81} Ampicillin-sulbactam, particularly the sulbactam component, and polymyxins have been used to treat carbapenem-resistant infections with some success. As with ESBL-producing Enterobacteriaceae, use of inappropriate empiric antibiotics has been associated with in-hospital mortality in the setting of serious infections.⁸²

Rates of clinical isolation of multidrug-resistant gram-negative bacilli vary between and within healthcare facilities. Knowledge of hospital-specific flora and antibiotic susceptibility patterns is crucial in the management, prevention, and control of serious healthcare-associated gram-negative infections.

CLOSTRIDIUM DIFFICILE INFECTION

C. difficile infection (CDI) remains a significant problem in healthcare facilities and is the number 1 cause of healthcare-associated diarrhea in the United

States. *C. difficile* is an anaerobic, gram-positive, spore-forming rod that can live in the hospital environment in a spore state for prolonged periods of time.⁸³ Nosocomial transmission occurs by the fecal-oral route. Susceptible hosts are usually those with altered natural colonic flora most often caused by antibiotics or chemotherapeutic agents.^{84,85} Acid-resistant *C. difficile* spores are able to pass through the stomach into the bowels, where they germinate, proliferate, and release toxins A and B (which have enterotoxin and cytotoxin activity). Both toxins adhere to receptors on the colonic brush border and cause necrosis and shedding of these cells into the lumen, leaving a shallow ulcer on the mucosal surface. Specifically, the toxins target Rho family proteins, causing dysregulation and actin filament disintegration, cytoskeleton disruption, and ultimately cell death.⁸⁶ The recently described, more virulent *C. difficile* strain, BI/NAP1, produces 16- to 23-fold more toxin A and toxin B than historical control strains.⁸⁷ The increased toxin production in vitro is thought to be secondary to a partial deletion within the organism's *tcdC* gene, which normally acts as a negative regulator of toxin A and toxin B production.⁸⁸ There may also be other mechanisms for increased virulence of this strain.

The hospital environment and hospitalized patients are the major reservoirs of *C. difficile*. The duration of hospitalization has been shown to correlate with the acquisition of CDI. Additionally, exposure to antimicrobials of any kind, especially to third-generation cephalosporins, has been implicated as a risk for CDI.⁸⁹ The newer BI/NAP1 strain is resistant to the fluoroquinolones, and infection with this strain is linked specifically to fluoroquinolone exposure and increased use of fluoroquinolones in certain institutions.⁹⁰ Other notable risk factors for infection include age over 65 years and any interventions or agents that alter the usual gastrointestinal flora or structure, including proton pump inhibitors, chemotherapeutic agents, nasogastric tubes, and gastrointestinal surgery.⁹¹

Clinical symptoms of *C. difficile* colitis usually begin 5 to 10 days after initiation of antibiotic therapy but can occur as early as 1 day after initiation or as late as 10 weeks after antibiotics are stopped.⁹² Symptoms can range in severity from mild, watery stools a few times per day to a systemic illness with fever, malaise, abdominal pain, nausea, and vomiting accompanied by many watery stools per day and signs of dehydration. Severe CDIs can lead to toxic megacolon, sepsis, and death.

Pathologically, gross findings in a colon with CDI can span the spectrum from normal to mild hyperemia to scattered ulcerations to diffuse

pseudomembrane formation. Microscopically, the pseudomembrane is composed of an inflammatory exudate composed of serum proteins, mucus, and dead cellular material.

The gold standard for the diagnosis of CDI is the cell culture cytotoxicity assay. This test has an excellent sensitivity but requires 24 to 48 hours for results and is both labor-intensive and expensive. Consequently, most hospitals rely on enzyme-linked immunoabsorbent assays to detect the toxin.⁹³ These tests are easier, more rapid, and less expensive, but they are also less sensitive than cytotoxic assays, with sensitivities ranging from 79 to 97%.⁹⁴ Polymerase chain reaction testing for *C. difficile* toxin genes has shown promise in several recent studies, but at this time, it remains a research tool.⁹⁵ Lastly, colonoscopic documentation of pseudomembranes is usually reserved for those cases with negative enzyme immunoassay toxin assays but a high clinical index of suspicion.

Treatment of CDI includes the discontinuation of the offending agents, avoidance of antiperistaltic agents, and usually either oral metronidazole or vancomycin. The updated Society for Healthcare Epidemiology of America/Infectious Diseases Society of America guidelines are expected to stratify patients into 3 categories of disease (Table 2).

Metronidazole is recommended for mild to moderate disease. Oral vancomycin is preferred for patients with more severe disease. For patients with severe disease complicated by ICU admission, toxic megacolon, or the need for a colectomy, high-dose vancomycin by a nasogastric tube and intravenous metronidazole are recommended. Multiple recurrences of CDI should be managed with an oral vancomycin taper, with or without pulse dosing.⁹⁶ Use of intravenous immunoglobulin and vancomycin enemas are controversial, and success has been anecdotal.

The prevention and control of healthcare-associated CDI requires a joint approach of preventing horizontal transmission within an institution along with the practice of good antimicrobial stewardship. *C. difficile* spores can grossly contaminate the hospital environment and are somewhat resistant to alcohol-based hand hygiene agents; thus, hand washing with soap and water is preferred after contact with a CDI patient. Ideally, infected patients should be placed in private rooms with private bathrooms. Staff members entering the patient's room are expected to wear gowns and gloves. A 1:10 bleach solution should be used to terminally clean an infected patient's room.⁹⁷

Table 2. Society for Healthcare Epidemiology of America/Infectious Diseases Society of America Recommendations for the Treatment of *Clostridium difficile* Infection Stratified by Disease Incidence and Severity.

Clinical Definition	Supportive Clinical Data	Recommended Treatment
Initial episode, mild or moderate	Leukocytosis <15,000 cells/mm ³ and creatinine <1.5× level prior to <i>C. difficile</i> infection	Oral metronidazole, 500 mg 3 times a day for 10–14 days
Initial episode, severe	Leukocytosis ≥15,000 cells/mm ³ or creatinine ≥1.5× level prior to <i>C. difficile</i> infection	Oral vancomycin, 125 mg 4 times a day for 10–14 days
Initial episode, complicated	Hypotension or shock, megacolon, perforation, and severe colitis seen on computed tomography scan	If no complete ileus: oral or nasogastric vancomycin, 500 mg 4 times a day, and/or intravenous metronidazole, 500–750 mg every 8 hours If complete ileus: intravenous metronidazole and rectal vancomycin
First recurrence		Same as for initial episode
Second or further recurrence		Oral vancomycin taper with or without pulse dosing: 125 mg orally 4 times daily for 7 days 125 mg orally twice daily for 7 days 125 mg orally once a day for 7 days 125 mg orally every other day for 7 days 125 mg orally every 3 days for 14 days

NOTE: This table was adapted from Cohen et al.¹⁰² The final version of the guidelines may differ from what is presented here.

Appropriate antibiotic management can also help to prevent and control CDI. Minimizing antibiotic use and its duration, avoiding the use of broad-spectrum antibiotics whenever possible, and avoiding the use of other agents that affect gastrointestinal motility are recommended. The prophylactic use of vancomycin or metronidazole has not been shown to be beneficial.⁸⁵

CDI continues to be a significant problem in the healthcare setting. This problem has been compounded by the new strain BI/NAP1, which has been associated with more severe disease. The diagnosis of CDI usually relies on enzyme immunoassay toxin assays in the setting of clinical symptoms. Metronidazole and vancomycin remain the mainstays of treatment, although newer therapies are under investigation. Sound infection control practices and antimicrobial stewardship are crucial in defending against the spread of *C. difficile* in the hospital environment.

CONCLUSIONS

There is vast room for improvement in the prevention of healthcare-associated infections, and several evidence-based methods are available. On a practitioner level, the most effective and basic principle of infection control is hand hygiene. This includes hand washing, use of alcohol-based products, and keeping short, clean nails.⁹⁸ Institutions can reduce healthcare-associated infections through the development of focused policies and projects. In New York City, for example, the Greater New York Hospital Association and the United Hospital Fund have partnered with local area hospitals to implement a CLABSI quality improvement project. The project incorporates surveillance with some of the system-based approaches outlined earlier to reduce CLABSIs.^{99,100} Although this project is specific to CLABSIs, others can be developed to reduce the morbidity and mortality associated with healthcare-associated infections. The Institute for Healthcare Improvement has initiated its "5 Million Lives" campaign to reduce harm from US healthcare providers, and its Web site provides a forum for individual institutions to discuss their experiences in implementing system-wide changes.¹⁰¹ Providers at all levels are instrumental in reducing nosocomial infections and should be included in policy development for institutions.

Familiarity with institution-specific microbiology and antibiotic susceptibility patterns is imperative for the prevention, recognition, and management of serious infections. The increasing isolation of highly

resistant bacteria and increasing rates of CDI highlight the need for aggressive antibiotic stewardship to restrict the use of broad-spectrum antibiotics to appropriate clinical situations. Antibiotic stewardship may also aid in the streamlining of antibiotic regimens once susceptibility data are available. Enlisting infectious disease consultants for assistance in the treatment of complicated infections and infections with multidrug-resistant organisms is encouraged.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Wald HL, Kramer AM. Nonpayment for harms resulting from medical care: catheter-associated urinary tract infections. *JAMA* 2007; 298: 2782–2784.
2. Kohn LT, Corrigan JM. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
3. Centers for Disease Control and Prevention. National Healthcare Safety Network. Available at: <http://www.cdc.gov/ncidod/dhqp/nhsn.html>. Accessed September 2008.
4. Darouiche RO. Device infections: a macroproblem that starts with microadherence. *Clin Infect Dis* 2001; 33: 1567–1572.
5. Centers for Disease Control and Prevention. *National Healthcare Safety Network Manual: Patient Safety Component Protocol*. Available at: www.cdc.gov/ncidod/dhqp/pdf/nhsn/nhsn_manual_patientsafetyprotocol_current.pdf. Accessed September 2008.
6. Mermel LA. Prevention of intravascular catheter-related infections. *Ann Intern Med* 2000; 132: 391–402.
7. Hana R, Raad IR. Diagnosis of catheter-related bloodstream infection. *Curr Infect Dis Rep* 2005; 7: 413–419.
8. Pittet D, Tarara D, Wenzel RP. Nosocomial bloodstream infection in critically ill patients. Excess length of stay, extra costs, and attributable mortality. *JAMA* 1994; 271: 1598–1601.
9. Maki DG, Kluger DM, Crnich CJ. The risk of bloodstream infection in adults with different intravascular devices: a systematic review of 200 published prospective studies. *Mayo Clin Proc* 2006; 81: 1159–1171.
10. Hugonnet S, Sax H, Eggimann P, et al. Nosocomial bloodstream infections and clinical sepsis. *Emerg Infect Dis* 2004; 10: 76–81.
11. Krein SL, Hofer TP, Kowalski CP, et al. Use of central venous catheter-related bloodstream infection prevention practices by US hospitals. *Mayo Clin Proc* 2007; 82: 672–678.
12. O'Grady NP, Alexander M, Dellinger EP, et al. Guidelines for the prevention of intravascular catheter-related infections. *MMWR Recomm Rep* 2002; 51: 1–29.

13. Institute for Healthcare Improvement. Implement the central line bundle. Available at: <http://www.ihl.org/ihl/topics/criticalcare/intensivecare/changes/implementthecentrallinebundle.htm>. Accessed September 2008.
14. Bleasdale SC, Trick WE, Gonzalez IM, et al. Effectiveness of chlorhexidine bathing to reduce catheter-associated bloodstream infections in medical intensive care unit patients. *Arch Intern Med* 2007; 167: 2073–2079.
15. Maki DG, Mermel LA, Klugar D, et al. The efficacy of a chlorhexidine impregnated sponge (Biopatch) for the prevention of intravascular catheter-related infection—a prospective randomized controlled multicenter study [abstract]. Paper presented at: Interscience Conference on Antimicrobial Agents and Chemotherapy; 2000; Toronto, Canada, Accessed July 2008.
16. Niël-Weise BS, Stijnen T, van den Broek PJ. Anti-infective-treated central venous catheters: a systematic review of randomized controlled trials. *Intensive Care Med* 2007; 33: 2058–2068.
17. O'Grady NP, Alexander M, Dellinger EP, et al. Guidelines for the prevention of intravascular catheter-related infections. *MMWR* 2002; 10: 1–29.
18. Nicolle LE. The prevention of hospital-acquired urinary tract infection. *Clin Infect Dis* 2008; 46: 251–253.
19. Mandell G, Bennett J, Dolin R. *Principles and Practice of Infectious Diseases*. 6th ed. New York, NY: Churchill Livingstone; 2005.
20. Saint S, Kowalski CP, Kaufman SR, et al. Preventing hospital-acquired urinary tract infection in the United States: a national study. *Clin Infect Dis* 2008; 46: 243–250.
21. Saint S, Kaufman SR, Rogers MAM, et al. Condom versus indwelling urinary catheters: a randomized trial. *J Am Geriatr Soc* 2006; 54: 1055–1061.
22. Klevens RM, Morrison MA, Nadle J, et al. Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *JAMA* 2007; 298: 1763–1771.
23. Jarvis WR, Schlosser J, Chinn RY, Tweeten S. National prevalence of methicillin-resistant *Staphylococcus aureus* in inpatients at US health care facilities, 2006. *Am J Infect Control* 2007; 35: 631–637.
24. Seybold U, Kourbatova EV, Johnson JG, et al. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health care-associated blood stream infections. *Clin Infect Dis* 2006; 42: 647–656.
25. Centers for Disease Control and Prevention. Four pediatric deaths from community-acquired methicillin-resistant *Staphylococcus aureus*—Minnesota and North Dakota, 1997–1999. *MMWR* 1999; 48: 707–710.
26. Kollef MH. Limitations of vancomycin in the management of resistant staphylococcal infections. *Clin Infect Dis* 2007; 45(suppl 3): S191–S195.
27. Kuehnert MJ, Kruszon-Moran D, Hill HA, et al. Prevalence of *Staphylococcus aureus* nasal colonization in the United States, 2001–2002. *J Infect Dis* 2006; 193: 172–179.
28. Lowy FD. *Staphylococcal aureus* infections. *N Engl J Med* 1998; 339: 520–532.
29. Applebaum P. Microbiology of antibiotic resistance in *Staphylococcus aureus*. *Clin Infect Dis* 2007; 45(suppl 3): S165–S170.
30. Harbarth S, Fankhauser C, Schrenzel J, et al. Universal screening for methicillin-resistant *Staphylococcal aureus* at hospital admission and nosocomial infection in surgical patients. *JAMA* 2008; 299: 1149–1157.
31. Robicsek A, Beaumont J, Paule S, et al. Universal surveillance for methicillin-resistant *Staphylococcus aureus* in 3 affiliated hospitals. *Ann Intern Med* 2008; 148: 409–418.
32. Everett ED, McNitt TR, Rahm AE, et al. Epidemiologic investigation of methicillin resistant *Staphylococcus aureus* in a burn unit. *Mil Med* 1978; 143: 165–167.
33. Klein E, Smith DL, Laxminarayan R. Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*, United States, 1999–2005. *Emerg Infect Dis* 2007; 13: 1840–1846.
34. Klevens RM, Edwards JR, Tenover FC, et al. Changes in the epidemiology of *Staphylococcus aureus* in intensive care units in US hospitals, 1992–2003. *Clin Infect Dis* 2006; 42: 389–391.
35. Oztoprak N, Cevik MA, Akinci EA, et al. Risk factors for ICU-acquired methicillin-resistant *Staphylococcus aureus* infections. *Am J Infect Control* 2006; 34: 1–5.
36. Tacconelli E, De Angelis G, Cataldo MA, et al. Does antibiotic exposure increase the risk of methicillin-resistant *Staphylococcus aureus* (MRSA) isolation? A systematic review and meta-analysis. *J Antimicrob Chemother* 2008; 61: 26–38.
37. Jernigan JA, Arnold K, Heilpern K, et al. Methicillin-resistant *Staphylococcus aureus* as community pathogen [conference summary]. Available at: <http://www.cdc.gov/ncidod/eid/vol12no11/06-0911.htm>. Accessed September 2008.
38. King MD, Humphrey BJ, Wang YF, et al. Emergence of community-acquired methicillin-resistant *Staphylococcus aureus* USA 300 clone as the predominant cause of skin and soft-tissue infections. *Ann Intern Med* 2006; 144: 309–317.
39. Hamilton SM, Bryant AE, Carroll KC, et al. In vitro production of Pantón-Valentine leukocidin among strains of methicillin-resistant *Staphylococcus aureus* causing diverse infections. *Clin Infect Dis* 2007; 45: 1550–1558.
40. Diep BA, Chambers HF, Graber CJ, et al. Emergence of multidrug-resistant, community-associated, methicillin-resistant *Staphylococcus aureus* clone USA300 in men who have sex with men. *Ann Intern Med* 2008; 148: 249–257.
41. Moran GJ, Krishnadasan A, Gorwitz RJ, et al. Methicillin-resistant *S. aureus* infections among patients in the emergency department. *N Engl J Med* 2006; 355: 666–674.
42. Treatment of community-associated MRSA infections. *Med Lett* 2006; 48: 13–14.
43. Baddour LM, Wilson WR, Bayer AS, et al. Infective endocarditis: diagnosis, antimicrobial therapy, and management of complications. *Circulation* 2005; 111: e394–e433.
44. Soriano A, Marco F, Martínez JA, et al. Influence of vancomycin minimum inhibitory concentration on the treatment of methicillin-resistant *Staphylococcus aureus* bacteremia. *Clin Infect Dis* 2008; 46: 193–200.

45. Hidayat LK, Hsu DL, Quist R, et al. High-dose vancomycin therapy for methicillin-resistant *Staphylococcus aureus* infections. *Arch Intern Med* 2006; 166: 2138–2144.
46. Jeffres MN, Isakow W, Doherty JA, et al. Predictors of mortality for methicillin-resistant *Staphylococcus aureus* health-care-associated pneumonia. *Chest* 2006; 130: 947–955.
47. Liu C, Chambers HF. *Staphylococcus aureus* with heterogeneous resistance to vancomycin: epidemiology, clinical significance, and critical assessment of diagnostic methods. *Antimicrob Agents Chemother* 2003; 47: 3040–3045.
48. Fowler VG, Boucher HW, Corey R, et al. Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N Engl J Med* 2006; 355: 653–665.
49. Wunderink RG, Rello J, Cammarata SK, et al. Linezolid vs vancomycin. *Chest* 2002; 124: 1789–1797.
50. Muto CA, Jernigan JA, Ostrowsky BE, et al. SHEA guideline for preventing nosocomial transmission of multidrug-resistant strains of *Staphylococcus aureus* and *Enterococcus*. *Infect Control Hosp Epidemiol* 2003; 24: 362–386.
51. Huang SS, Yokoe DS, Hinrichsen VL, et al. Impact of routine intensive care unit surveillance cultures and resultant barrier precautions on hospital-wide methicillin-resistant *Staphylococcus aureus* bacteremia. *Clin Infect Dis* 2006; 43: 971–978.
52. Lautenbach E. Expanding the universe of methicillin-resistant *Staphylococcus aureus* prevention. *Ann Intern Med* 2008; 148: 474–476.
53. Gaynes R, Edwards JR, and the National Nosocomial Infections Surveillance System. Overview of nosocomial infections caused by gram-negative bacilli. *Clin Infect Dis* 2005; 41: 848–854.
54. Philippon A, Labia R, Jacoby G. Extended-spectrum β -lactamases. *Antimicrob Agents Chemother* 1989; 33: 1131–1136.
55. Jacoby GA, Medeiros AA, O'Brien TF, et al. Broad-spectrum transmissible β -lactamases. *N Engl J Med* 1988; 319: 723–724.
56. Paterson DL, Ko WC, Gottberg AV, et al. International prospective of *Klebsiella pneumoniae* bacteremia: implications of ESBL production in nosocomial infections. *Arch Intern Med* 2004; 140: 26–32.
57. National Nosocomial Infections Surveillance (NNIS) system report, data summary from January 1992 through June 2004, issued October 2004. *Am J Infect Control* 2004; 32: 470–485.
58. Lautenbach E, Patel JB, Bilker WB, et al. Extended-spectrum β -lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*: risk factors for infection and impact of resistance on outcomes. *Clin Infect Dis* 2001; 32: 1162–1171.
59. Jacoby GA, Munoz-Price LS. The new β -lactamases. *N Engl J Med* 2005; 352: 380–391.
60. Hyle EP, Lipworth AD, Zaoutis TE, et al. Risk factors for increasing multi-drug resistance among extended-spectrum β -lactamase producing *Escherichia coli* and *Klebsiella* species. *Clin Infect Dis* 2005; 40: 1317–1324.
61. Paterson DL, Mulazimoglu L, Casellas JM, et al. Epidemiology of ciprofloxacin resistance and its relationship to extended-spectrum β -lactamase production in *Klebsiella pneumoniae* isolates producing bacteremia. *Clin Infect Dis* 2000; 30: 473–478.
62. Paterson DL. Recommendations for treatment of severe infections caused by Enterobacteriaceae producing extended-spectrum β -lactamases (ESBLs). *Clin Microbiol Infect* 2000; 6: 460–463.
63. Meyer KS, Urban C, Eagan JA, et al. Nosocomial outbreak of *Klebsiella* infection resistant to third-generation cephalosporins. *Ann Intern Med* 1993; 119: 353–358.
64. Karas JA, Pillay DG, Muckart D, Sturm AW. Treatment failure due to extended-spectrum β -lactamase. *J Antimicrob Chemother* 1996; 37: 203–204.
65. Paterson DL, Ko W, Gottberg A, et al. Outcome of cephalosporin treatment for serious infections due to apparently susceptible organisms producing extended-spectrum β -lactamases: implications for the microbiology laboratory. *J Clin Microbiol* 2001; 39: 2206–2212.
66. Paterson DL, Ko W, Gottberg A, et al. In vitro susceptibility and clinical outcomes of bacteremia due to extended-spectrum β -lactamase producing *Klebsiella pneumoniae*. *Clin Infect Dis* 1998; 27: 956.
67. Hyle EP, Lipworth AD, Zaoutis TE, et al. Impact of inadequate initial antimicrobial therapy on mortality in infections due to extended-spectrum β -lactamase-producing Enterobacteriaceae: variability by site of infection. *Arch Intern Med* 2005; 165: 1375–1380.
68. Kollef MH, Sherman G, Ward S. Inadequate antimicrobial treatment of infections: a risk factor for hospital mortality among critically ill patients. *Chest* 1999; 115: 462–474.
69. Harris AD, Smith D, Johnson JA, et al. Risk factors for imipenem-resistant *Pseudomonas aeruginosa* among hospitalized patients. *Clin Infect Dis* 2002; 34: 340–345.
70. Abbo A, Carmeli Y, Navon-Venezia S, et al. Impact of multi-drug-resistant *Acinetobacter baumannii* on clinical outcomes. *Eur J Clin Microbiol Infect Dis* 2007; 26: 793–800.
71. Lee S, Kim NJ, Choi S, et al. Risk factors for acquisition of imipenem-resistant *Acinetobacter baumannii*: a case-control study. *Antimicrob Agents Chemother* 2004; 48: 224–228.
72. Ahmad M, Urban C, Mariano N, et al. Clinical characteristics and molecular epidemiology associated with imipenem-resistant *Klebsiella pneumoniae*. *Clin Infect Dis* 1999; 29: 352–355.
73. Bratu S, Tolaney P, Karumudi U, et al. Carbapenemase-producing *Klebsiella pneumoniae* in Brooklyn, NY: molecular epidemiology and in vitro activity of polymyxin B and other agents. *J Antimicrob Chemother* 2005; 56: 128–132.
74. Patel G, Huprikar S, Jenkins SG, Calfee DP. Carbapenem-resistant *Klebsiella pneumoniae*: risk factors and mortality [abstract 563]. Paper presented at: 45th Annual Meeting of the Infectious Diseases Society of America; October 2007; San Diego, CA.
75. Villegas MV, Hartstein AI. *Acinetobacter* outbreaks. *Infect Control Hosp Epidemiol* 2003; 24: 284–295.
76. Paterson DL. The epidemiologic profile of infections with multidrug-resistant *Pseudomonas aeruginosa* and *Acinetobacter* species. *Clin Infect Dis* 2006; 43: S43–S48.
77. Munoz-Price LS, Weinstein RA. *Acinetobacter* infection. *N Engl J Med* 2008; 358: 1271–1281.

78. Bonomo RA, Szabo D. Mechanisms of multidrug resistance in *Acinetobacter* species and *Pseudomonas aeruginosa*. *Clin Infect Dis* 2006; 43: S49–S56.
79. Go ES, Urban C, Burns J, et al. Clinical and molecular epidemiology of *Acinetobacter* infections sensitive only to polymyxin B and sulbactam. *Lancet* 1994; 344: 1329–1333.
80. Garnacho-Montero J, Ortiz-Leyba C, Fernandez-Hinojosa E, et al. *Acinetobacter baumannii* ventilator-associated pneumonia: epidemiological and clinical findings. *Intensive Care Med* 2005; 31: 549–555.
81. Landman DL, Bratu S, Kochar S, et al. Evolution of antimicrobial resistance among *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, and *Klebsiella pneumoniae* in Brooklyn, NY. *J Antimicrob Chemother* 2007; 20: 78–82.
82. Falagas ME, Kasiakou SK, Rafailidis PI, et al. Comparison of mortality of patients with *Acinetobacter baumannii* bacteraemia receiving appropriate and inappropriate empirical therapy. *J Antimicrob Chemother* 2006; 57: 1251–1254.
83. Bartlett JG. Narrative review: the new epidemic of *Clostridium difficile* associated enteric disease. *Ann Intern Med* 2006; 145: 758–764.
84. Robertson MB, Breen LJ, Desmond PV, et al. Incidence of antibiotic related diarrhea and pseudomembranous colitis: a prospective study of lincomycin, clindamycin and ampicillin. *Med J Aust* 1977; 1: 243–246.
85. Anand A, Glatt AE. *Clostridium difficile* infection associated with antineoplastic chemotherapy: a review. *Clin Infect Dis* 1993; 17: 109–113.
86. Hippenstiel J, Kratz T, Krull M, et al. Rho protein inhibition blocks protein kinase C translocation and activation. *Biochem Biophys Res Commun* 1998; 245: 830–834.
87. Loo V, Poirier L, Miller MA, et al. A predominantly clonal multi-institutional outbreak of *Clostridium difficile* associated diarrhea with high morbidity and mortality. *N Engl J Med* 2005; 353: 2442–2449.
88. Warny M, Pepin J, Fang A, et al. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* 2005; 366: 1079–1084.
89. Johnson S, Gerding DN. *Clostridium difficile* associated diarrhea. *Clin Infect Dis* 1998; 26: 1027–1036.
90. Muto CA, Pokrywka M, Shutt K, et al. A large outbreak of *Clostridium difficile* associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. *Infect Control Hosp Epidemiol* 2005; 26: 273–280.
91. Bignardi GE. Risk factors for *Clostridium difficile* infection. *J Hosp Infect* 1998; 40: 1–15.
92. Tedesco FJ. Pseudomembranous colitis: pathogenesis and therapy. *Med Clin North Am* 1982; 66: 655–664.
93. Gerding DN, Johnson S, Peterson LR, et al. *Clostridium difficile* associated diarrhea and colitis. *Infect Control Hosp Epidemiol* 1995; 16: 459–477.
94. Manabe YC, Vinetz JM, Moore RD, et al. *Clostridium difficile* colitis: an efficient clinical approach to diagnosis. *Ann Intern Med* 1995; 123: 835–840.
95. Belanger S, Boissinot M, Clairoux N, et al. Rapid detection of *Clostridium difficile* in feces by real time PCR. *J Clin Microbiol* 2003; 41: 730–734.
96. Tedesco FJ, Gordon D, Fortson WC. Approach to patients with multiple relapses of antibiotic-associated pseudomembranous colitis. *Am J Gastroenterol* 1985; 80: 867.
97. Boyce JM, Pittet D, for the Healthcare Infection Control Practice Advisory Committee, Society for Healthcare Epidemiology in America, Association for Professionals in Infection Control, and Infectious Diseases Society of America. Hand Hygiene Task Force guideline for hand hygiene in healthcare settings: recommendations of the Healthcare Infection Control Practice Advisory Committee and HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. *Infect Control Hosp Epidemiol* 2002; 23(suppl): S3–S40.
98. Siegel JD, Rhinehard E, Jackson M, Chiarello L, and the Healthcare Infection Control Practices Advisory Committee. Guideline for isolation precautions: preventing transmission of infections agents in healthcare settings, 2007. Available at: <http://www.cdc.gov/ncidod/dhqp/pdf/guidelines/isolation2007.pdf>. Accessed September 2008.
99. United Hospital Fund. New York hospitals join forces to target infections in ICUs. Available at: http://www.uhfnyc.org/press_release3159/press_release_show.htm?doc_id=367431. Accessed September 2008.
100. Ortega D, Chen BYL, Rahman M, et al. A multidisciplinary approach to the prevention and control of central line-associated bloodstream infections (CLABS) in the intensive care unit (ICU) of a community hospital in Central Harlem, New York. *Am J Infect Control* 2007; 35: E153–E154.
101. Institute for Healthcare Improvement. Protecting 5 million lives from harm. Available at: <http://www.ihf.org/ihf/programs/campaign>. Accessed July 2008.
102. Cohen SH, Gerding D, Johnson S, et al. IDSA proposed guidelines for the treatment of *C. difficile* diarrhea. Paper presented at: 45th Annual Meeting of the Infectious Diseases Society of America; October 4–7, 2007 San Diego, CA.

**HINARI PHASE I COUNTRIES**

Free access to more than 100 biomedical journals published by Wiley

**HINARI PHASE II COUNTRIES**

Nominal cost access to more than 100 biomedical journals published by Wiley

Developing World Access to Research

“perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries.”

Former WHO Director-General Dr Gro Harlem Brundtland (July, 2001)

Wiley, together with the World Health Organisation and fellow Scientific, Technical and Medical publishers, is proud to play a role in furthering global dissemination of biomedical information.

Through the Health InterNetwork Access to Research Initiative (HINARI) www.healthinternetwork.org, free or nominal cost online access to more than 2,200 biomedical journals from 42 publishers is provided to more than 1,000 institutions in approximately 100 of the world's lowest income countries.

Wiley is one of the six founding publishers who led this initiative with the WHO. We offer more than 100 of our leading biomedical journals through the HINARI program; including many leading journals published with our society partners.

HINARI
HEALTH INTERNETWORK
ACCESS TO RESEARCH

**WILEY****InterScience®**

DISCOVER SOMETHING GREAT

www.interscience.wiley.com

Wiley – helping to make a world of difference.

Giant
Resources...
Purebred
Content

Wiley InterScience® Reference Works

Sample Title List

Encyclopedia of Life Sciences
Encyclopedia of Statistical Sciences
Kirk-Othmer Encyclopedia of Chemical Technology
Sax's Dangerous Properties of Industrial Materials
Ullmann's Encyclopedia of Industrial Chemistry
Van Nostrand's Encyclopedia of Chemistry

Capabilities

Full text HTML and PDF (for some)
Reference linking
Citation download

Access

Single article or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Today's sophisticated researcher demands fast and convenient online access to the giants of scientific reference.

From *Kirk-Othmer* to the *Encyclopedia of Life Sciences*,

Wiley InterScience® Reference Works deliver the biggest names in scientific research literature to your desktop, right when you need them. Encompassing subjects ranging from chemistry, life sciences, and medicine to psychology, engineering, and physics, these multi-volume encyclopedias, handbooks, and dictionaries are enhanced by powerful functionality that accelerates your path to discovery. With **Wiley InterScience**, you can be sure to **Discover Something Great.**



WILEY

InterScience®

DISCOVER SOMETHING GREAT

 **WILEY**

1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 76, NO. 2
MARCH/APRIL 2009

Issue Theme
Traumatic Brain Injury:
From Bench to
Bedside to Home

Theme Editors
Silvana Riggio, MD
Andy Jagoda, MD

Journal of Translational and Personalized Medicine

**Neurophysiology of Traumatic
Brain Injury**

**Epidemiology of Traumatic
Brain Injury in United States**

**Post-traumatic Traumatic Brain
Injury: Mechanisms of Injury**

**Severe Traumatic Brain Injury:
Maximizing Outcomes**

Child Traumatic Brain Injury

**Prehospital Management:
First Link in Chain of Survival**

**Neuroimaging of
Traumatic Brain Injury**

**Neurobehavioral Sequelae of
Traumatic Brain Injury**

**Cognitive Rehabilitation:
Assessment to Treatment**

**Neurorehabilitation:
Does it Make a Difference?**

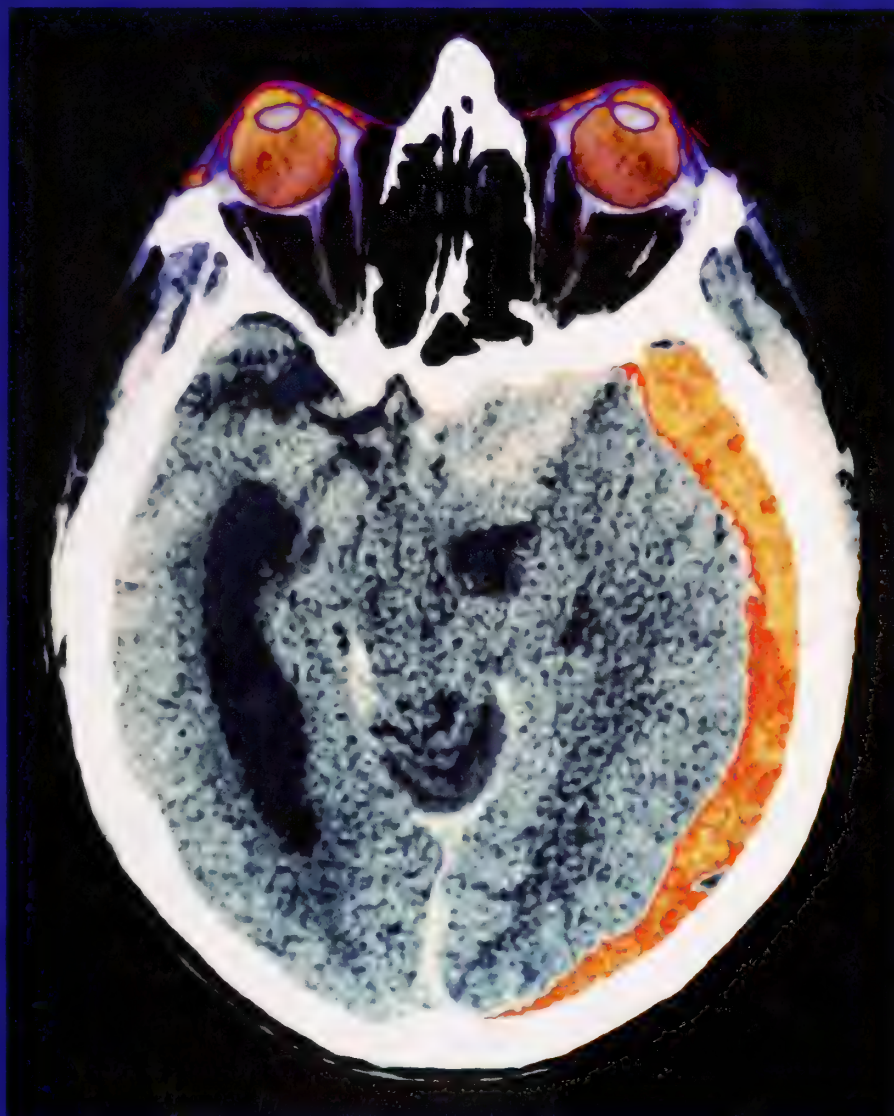
**Prognostic
Value of
Traumatic Brain Injury**

Special Features

Cardiovascular Safety of Medications for Attention-deficit/Hyperactivity Disorder

Case Study: Clopidogrel for Preventing Cardiovascular Events

Physical Exercise and Depression



Newly
Relaunched!



Keep Abreast of the Latest Trends in
Translational and Personalized Medicine...
Enter Your Subscription Today!

MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Editor-in-Chief:

Penny A. Asbell, MD, FACS, MBA

Mount Sinai School of Medicine, New York, NY

2008 Themes and Guest Editors

Quality of Healthcare and Racial
Health Disparities (Mark Chassin)

9/11 World Trade Center Disaster:
Past and Future (Robin Herbert,
Jacqueline Moline)

Anxiety and Depression:
Leading Edge of Therapy
(Dennis Charney, Mathew Sanjay)

Diabetes Type 1:
New Light on Etiology,
New Approaches to Management
(Derek Leroith)

Hospital Medicine:
An Ascending New Field
(Andrew Dunn)

Advanced Lung Disease:
Meeting the Challenges
(Maria Padilla)

Continuously published since 1934 and included in *Index Medicus* from the very first volume, the **Mount Sinai Journal of Medicine (MSJM)** has a distinguished career as a general medical journal for a worldwide audience.

Building upon this tradition of education across medical specialties, **MSJM** was re-launched in 2007 as a journal of translational and personalized medicine, two disciplines in the forefront of modern healthcare. **MSJM** features articles on the evolving nature of clinical care, its close linkage to bio-medical research and the resulting opportunities for enhanced patient care.

Dedicated to improving communication among all participants in medicine today, **MSJM** is an excellent source of up-to-date medical information for healthcare clinicians, educators, researchers and policy makers worldwide. **MSJM** will inspire the creative thinking needed to create new ways to understand, prevent, and treat human disease.

Each issue includes several quality, in-depth review articles related to a single theme and several others on subjects of diverse clinical interest. All articles are written by renowned authorities in their fields, from both Mount Sinai and elsewhere, under the supervision of Guest Editors.

 **WILEY-BLACKWELL**

Enter Your 2009 Subscription Today!



Your 2009 subscription will include the six 2009 print issues of the **Mount Sinai Journal of Medicine** (Volume 76), plus concurrent full-text online access to these six issues as they are posted on Wiley's Internet publishing platform.

Two Ways for Individuals to Subscribe:

1. Mount Sinai Alumna/Alumnus:
\$40.00 (Worldwide)
2. Personal Subscription:
\$95.00 (U.S./Canada/Mexico)
\$120.00 (Outside North America)

Complimentary Online Access for Institutions!

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

PERSONAL SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 personal subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$95.00 (U.S./Canada/Mexico) ☐ \$120.00 (Outside North America)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

Method of Payment (Canada residents please add GST. Personal rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **John Wiley & Sons, Inc.** Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA

Card # _____ Exp. Date _____

Signature _____

(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM0909

MSJM MEMBER SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 Alumna/Alumnus Subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$40.00 (Worldwide)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

METHOD OF PAYMENT (Alumna/Alumnus rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **The Mount Sinai School of Medicine**. Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA ☐ Other

Card # _____ Exp. Date _____

Signature _____

(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM09

COMPLIMENTARY ONLINE ACCESS FOR INSTITUTIONS

Mount Sinai Journal of Medicine

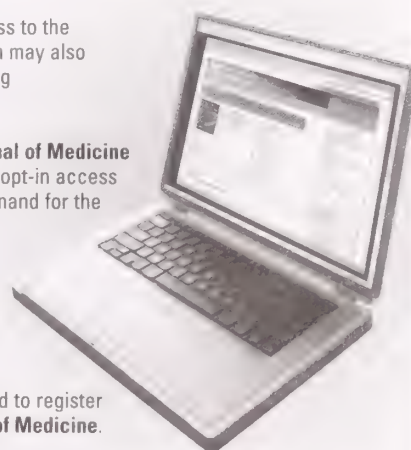
A Journal of Translational and Personalized Medicine

2009. Volume 76. 6 issues per year. Print ISSN: 0027-2507. Online ISSN: 1931-7581.

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

Institutional online and print pricing for the **Mount Sinai Journal of Medicine** will be set in 2009, subsequent to introductory complimentary opt-in access during 2008 and 2009. Pricing will be based on the overall demand for the Journal and on the size and type of the subscribing institution (e.g. academic, government, hospital, or corporate). Institutions that receive complimentary online access during 2008 and 2009 will have archival access in subsequent years according to Wiley's licensing terms and conditions.

Ask your librarian to e-mail optinaccess@wiley.co.uk or visit www.interscience.wiley.com/msjregister for more details and to register for complimentary online access to the **Mount Sinai Journal of Medicine**.





NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS MAIL

PERMIT NO. 2277

HOBOKEN, NJ

POSTAGE WILL BE PAID BY ADDRESSEE

JOHN WILEY & SONS INC
ATTN: PETER JANSIC, MAIL STOP 8-02
111 RIVER STREET
HOBOKEN, NJ 07030-9441



PLACE
POSTAGE
STAMP
HERE

INES SANTOS
MOUNT SINAI JOURNAL OF MEDICINE
ONE GUSTAV L. LEVY PLACE BOX 1094
NEW YORK, NY 10029-6574

**Library
Routing
Card**

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis S. Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner
Thomas M. Coffman
Dwight Evans
Gabriel Haddad
Heidi E. Hamm
Matti Häyry
Eve J. Higginbotham
Katherine A. High
Tony Hope
John H. Krystal
Raju Kucherlapati
Robert J. Lefkowitz
Robert Lisak
George Lister
Philip A. Marsden
Charles Nemeroff
Scott Rauch
Bernard Roizman
Alan F. Schatzberg
Inder Verma
Jeffrey A. Whitsett
Allan Wolkoff
Noam Zohar

University of Washington
Duke University
University of Pennsylvania
University of California, San Diego
Vanderbilt University
University of Manchester
Morehouse School of Medicine
University of Pennsylvania
University of Oxford
Yale University
Harvard Medical School
Duke University
Wayne State University
University of Texas
University of Toronto
Emory University
Harvard Medical School
University of Chicago
Stanford University
The Salk Institute
University of Cincinnati
Albert Einstein College of Medicine
Bar-Ilan University

Theme Editors

Silvana Riggio, MD
Associate Professor of
Psychiatry and Neurology
Mount Sinai School of Medicine

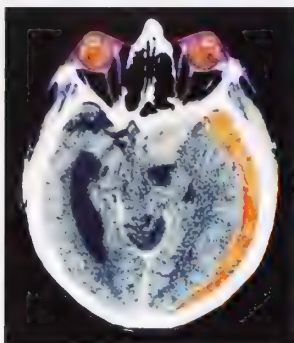
Andy Jagoda, MD, FACEP
Professor and Vice Chair of
Emergency Medicine
Mount Sinai School of Medicine

Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com



Traumatic Brain Injury

Coloured computed tomography (CT) scan of an axial section through a head showing a subdural hematoma (orange, right). Subdural hematoma occurs when an injury causes bleeding between the membranes that surround the brain. The blood rapidly clots to form a solid mass. Subdural haematoma is a dangerous condition, usually caused by impact to the head. Credit: Zephyr/Science Photo Library.

Themes for future issues of *MSJM*, each shown with Guest Editor(s)

- Type 2 diabetes mellitus (Dina Green)
- Zoonotic infections (Stephanie Factor)
- Medical education (Lisa Coplit)
- Alzheimer's disease (Nikos Robakis)
- Drug discovery in academic setting (Ravi Iyengar)
- Vascular surgery (Michael Marin)
- Parkinson's disease (Michele Tagliatti)

©2009 Mount Sinai School of Medicine. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: \$95.00 in North America, \$120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinajournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS

VOL 76, NO 2

Theme Introduction

- 95 Traumatic Brain Injury: From Bench to Bedside to Home
Silvana Riggio, Andy Jagoda

Theme Articles

- 97 Pathophysiology of Traumatic Brain Injury
Mark W. Greve, Brian J. Zink
- 105 Traumatic Brain Injury in the United States: An Epidemiological Overview
Carl R. Summers, Brian Ivins, Karen A. Schwab
- 111 Blast-related Traumatic Brain Injury: Mechanisms of Injury and Impact on Care
Gregory Elder, Adrian Cristian
- 119 Severe Traumatic Brain Injury: Maximizing Outcomes
Mary E. Tang, Darlene A. Lobel
- 129 Mild Traumatic Brain Injury
John J. Bruns, Jr, Andy S. Jagoda
- 138 Prehospital Management: First Link in Chain of Survival
Joseph Minardi, Todd J. Crocco
- 145 Neuroimaging of Traumatic Brain Injury
Tuong H. Le, Alisa D. Gean
- 163 Neurobehavioral Sequelae of Traumatic Brain Injury
Silvana Riggio, Meredith Wang
- 173 Cognitive Rehabilitation: Assessment to Treatment
Theodore Tsaousides, Wayne A. Gordon
- 182 Neurorehabilitation: Does it Make a Difference?
Brian D. Greenwald, John L. Rigg
- 190 The Future of Traumatic Brain Injury
Jamshid Ghajar

Feature Departments

- 194 Evidence-based Medicine [Editor: Thomas McGinn]
Clopidogrel for Preventing Cardiovascular Events
James Stulman
- 198 New Drugs and the FDA [Editor: Jeffery H. Newcorn]
Cardiovascular Safety of Medication Treatments for Attention-deficit/Hyperactivity Disorder
Jeffrey H. Newcorn, Craig Donnelly

Special Feature

- 204 Physical Exercise and Depression
Marije aan het Rot, Katherine A. Collins, Heidi L. Fitterling



don't publish online in the dark...



let our open linking network
shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-

profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at www.crossref.org, and make sure all your electronically published work is CrossRef-enabled.



THEME INTRODUCTION

Traumatic Brain Injury: From Bench to Bedside to Home

Silvana Riggio, MD,^{1,2} and Andy Jagoda, MD, FACEP³

¹Department of Psychiatry, Mount Sinai School of Medicine, New York, NY

²Department of Neurology, Mount Sinai School of Medicine, New York, NY

³Department of Emergency Medicine, Mount Sinai School of Medicine, New York, NY

Traumatic brain injury (TBI) spans a wide spectrum of injuries that range from the minor and inconsequential to the severe and disabling. Good outcomes from TBI depend on a coordinated health-care system that ensures comprehensive management from onset through recovery. A multidisciplinary approach is required in TBI management. This issue of the *Mount Sinai Journal of Medicine* is dedicated to demonstrating the importance of recognizing the continuum of care which is needed in order to maximize outcomes after TBI. The success of each link in the chain of care in TBI is dependent on the intervention provided by the preceding and succeeding links. Indeed, when final outcomes are considered, no one component in the continuum of care can pretend to exist independently of the others.

Conceptually, this issue of the journal is divided into 3 sections. The first series of articles looks at the epidemiology and pathophysiology with a special focus on blast injury as it relates to TBI. Blast injury is of increasing importance because of the types of explosive devices encountered by our military in the recent wars in Iraq and Afghanistan. This section establishes the gravity of the public health problem confronting both developed and developing countries. In the United States, there are approximately 1.5 million emergency department visits per year for

TBI, 150,000 hospitalizations, and 50,000 deaths—this is only the tip of the iceberg. There are countless millions of “mild” brain injuries for which medical care is never sought but that place the patients at risk for sequelae. The incidence of TBI is growing at an alarming rate in developing countries, where motorized vehicles combined with improved roads are creating the perfect setup for catastrophe. Understanding the epidemiology of TBI allows public health policy makers to focus on areas in need of intervention, as seen with bicycle helmet laws and motor vehicle restraint systems. Understanding the pathophysiology of TBI holds promise for identifying physiological and pharmacological interventions that minimize the secondary injuries that account for much of the disability after injury. However, despite an improved understanding of the pathophysiological mechanisms, benefits from pharmacological treatments have been elusive. The last article in this section provides an update on neuroimaging and its applications and limitations in diagnosing TBI.

The second section of this issue of the journal addresses acute management. The article on prehospital care establishes the critical role played by emergency medical systems. Emergency medical systems are responsible for identifying those patients at risk of a significant injury and directing their care to an appropriate trauma center. In cases of severe TBI, the neurosurgeon and/or neurointensivists have little hope of success if the prehospital care has not maximized brain oxygenation and perfusion. Perhaps this article more than any other demonstrates the dependence of final outcomes on individual components of a system. The succeeding articles on mild TBI and on severe TBI address issues related to acute management. Both of these articles build off evidence-based guidelines that have been developed with well-defined methodologies; both articles also

Address Correspondence to:

Andy Jagoda

Department of Emergency
Medicine
Mount Sinai School of Medicine
New York, NY
Email: andy.jagoda@mssm.edu

demonstrate the value of evidence-based guidelines in maximizing resource utilization and improving patient outcomes.

The third section of this issue is focused on the sequelae of TBI. The articles examine neurobehavioral manifestations, neuropsychological assessments, and neurorehabilitation. All 3 of these articles establish the challenge of assessing and defining injury: they emphasize the need for a coordinated approach to the post-TBI patient with a diagnostic and therapeutic approach that is tailored to the individual. The literature related to these topics reflects a critical need for well-designed studies to help define best practices in caring for the post-TBI patient.

The concept for this issue of the journal is the result of the coeditors' many years of collaborative experience in caring for TBI patients: Silvana Riggio

is a neurologist and psychiatrist, and Andy Jagoda is an emergency medicine physician. The issue's vision is to provide a state-of-the-art overview of TBI, to emphasize the importance of multidisciplinary care, and to demonstrate how research done at the basic science level can ultimately affect the success of neuroresuscitation and neurorehabilitation. The contributors to this issue of the journal have recognized expertise in their respective areas, and by bringing their experience and perspectives together in one place, we hope that a comprehensive perspective can be provided to stimulate further research and promote quality care for patients with TBI.

DISCLOSURES

Potential conflict of interest: Nothing to report.

Pathophysiology of Traumatic Brain Injury

Mark W. Greve, MD and Brian J. Zink, MD

Department of Emergency Medicine, Warren Alpert School of Medicine, Brown University, Providence, RI

ABSTRACT

Traumatic brain injury is a major source of death and disability worldwide. Significant success has been achieved in improving short-term outcomes in severe traumatic brain injury victims; however, there are still great limitations in our ability to return severe traumatic brain injury victims to high levels of functioning. Primary brain injury, due to initial injury forces, causes tissue distortion and destruction in the early postinjury period. Clinical outcomes depend in large part on mediating the biomolecular and cellular changes that occur after the initial injury. These secondary injuries from traumatic brain injury lead to alterations in cell function and propagation of injury through processes such as depolarization, excitotoxicity, disruption of calcium homeostasis, free-radical generation, blood-brain barrier disruption, ischemic injury, edema formation, and intracranial hypertension. The best hope for improving outcome in traumatic brain injury patients is a better understanding of these processes and the development of therapies that can limit secondary brain injury. *Mt Sinai J Med* 76:97–104, 2009. © 2009 Mount Sinai School of Medicine

Key Words: pathophysiology, traumatic brain injury, primary brain injury, secondary brain injury.

Traumatic brain injury (TBI) continues to be a leading cause of death and disability worldwide.¹ The links to survival and improved outcome

start with prehospital providers and extend to the operative suite and the intensive care unit. Focused protocol-driven management, including early intubation, rapid advanced cardiac life support transport, prompt resuscitation, early computed tomography, immediate evacuation of mass lesions, and meticulous management in an intensive care unit setting, has reduced mortality from 50% in the 1970s to 17% in 2003.^{2,3} Even brief periods of hypotension or hypoxia can cause significant brain tissue death and dysfunction.⁴ There continues to be an important window of opportunity to improve outcomes for those with severe brain injuries through an understanding of the pathophysiological processes involved.

Although the forces initiating the primary injury generally take less than 100 milliseconds to occur, the resulting pathophysiological events are much more prolonged and progressive. To better understand these processes, TBI has been classified into primary injury and secondary injury. The primary injury is understood as the result of the mechanical forces acting on the injured brain. The secondary injury is due to the biomolecular and physiological changes that follow these insults. This delineation is somewhat artificial as the 2 processes are intimately linked. The biomolecular and physiological changes that occur following TBI are the subject of this article.

METHODS

Two searches were conducted with an Ovid search engine of Medline databases. The first search limited of English language reviews within the last 5 years on physiology or pathophysiology of TBI in human subjects yielded 27 articles. A second Ovid search of Medline under the parameters of TBI, English language, core clinical journals, clinical trials over the last 10 years, and age over 19 yielded 32 articles. We reviewed and used the guidelines of the Brain Trauma Foundation⁵ to provide a management basis

Address Correspondence to:

Mark W. Greve

Department of Emergency
Medicine

Warren Alpert School
of Medicine

Brown University
Providence, RI

Email: mgreve@lifespan.org

for focusing our discussion and deciding which articles to include in this review. Also included in this review are prior reviews of TBI by experts in the field.

BIOMECHANICS OF BRAIN INJURY

The initial injury can be not only the result of direct trauma but also the result of acceleration, deceleration, or rotational forces either concurrent with or independent of direct trauma. This cascade of events leads to inertial forces on the brain tissue and cells. Biomechanical theories have historically described 2 inertial phenomena: linear acceleration and rotational head movement. It is thought that linear acceleration forces lead to superficial brain lesions, whereas rotational movements may explain deeper cerebral lesions and the concussion mechanism.⁶ The tissue strains induced by both linear and rotational forces create spatiotemporal gradients.⁷ Because the brain is a viscoelastic organ with little internal structural support, it poorly tolerates such forces. The gray matter closest to the surface of the brain is most susceptible to linear forces, which may cause cortical contusions and hemorrhage.⁸ The deeper cerebral white matter axons can be physiologically and mechanically injured by rotational forces. This disruption of deeper white matter is called diffuse axonal injury.⁹ Deep gray matter nuclei and axonal tracts in the midbrain and brainstem may also be damaged by rotational forces.

The linear and rotational acceleration theories are considered by many to be incomplete and may not adequately explain injury to deeper cortical structures in the absence of injury to more superficial cerebral structures. Although largely theoretical in its application to TBI, the stereotactile theory may explain these injuries. The stereotactile theory considers the spherical shape of the cranial vault in the setting of skull-brain relative movements and skull vibrations and their ability to generate secondary pressure waves. As brain tissue generally has the same density on concentric planes, these waves may propagate as a spherical wave front. The spherical nature of this wave front and the pressure gradient propagates and focuses its energy on deeper cerebral structures.^{10,11}

Apart from the forces incurred from the initial trauma, the prolonged compressive forces of brain edema and hematomas can further impair brain function by distorting brain tissues, elevating intracranial pressure (ICP), and reducing cerebral

blood flow (CBF). Penetrating brain injury obviously harms brain tissues that are directly in the path of the projectile, but the resultant blast effect also produces significant structural damage to adjacent tissues. With penetrating injury, the trajectory and location of the wound are the most significant factors in outcome.

BIOMOLECULAR RESPONSE TO BRAIN INJURY

Compared to the primary injury, secondary injuries to the brain are more indolent and progressive and may ultimately be the deciding factors in the patient's recovery. The early primary injury period is characterized by the destruction of brain tissue but not widespread degeneration of neurons, glial cells, or axons; these later effects are in essence the results of secondary injury.¹² For patients who survive the initial injury, morbidity and mortality will be determined by secondary injury processes. Despite numerous efforts to pharmacologically intervene in secondary injury, none have demonstrated clinical efficacy. Clearly, there is ample opportunity to improve the care of these patients, and much of the promise lies at the cellular and subcellular level.

Substantial work in animal models, cell preparations, and human data have elucidated a number of biomolecular responses to TBI (see Figure 1). Excitatory amino acid release, oxygen radical reactions, and nitric oxide production are closely intertwined in the brain cell response to injury. One theory is that excitatory amino acids are released and the associated calcium influx into neurons and other brain cells is the sentinel event following TBI. Excess intracellular calcium then promotes oxygen radical reactions. High calcium and the presence of free-radical molecules create an unstable environment in the cell that may lead to increased production and release of nitric oxide and excitatory amino acids (eg, glutamate). Nitric oxide may participate in oxygen radical reactions and lipid peroxidation in neighboring cells, with a subsequent release of excitatory amino acids. Thus, each component of the triad—glutamate, free radicals, and nitric oxide—has the potential to enhance the activity of the others. If endogenous protective mechanisms such as free-radical scavengers do not halt this cycle, widespread cellular damage or necrosis will result.

The initial cellular response to injury is widespread and massive depolarization of neurons, glial cells, and cerebral vascular endothelial cells. Excess glutamate is thought to be the cause of the depolarization of traumatized cells in TBI, but there is

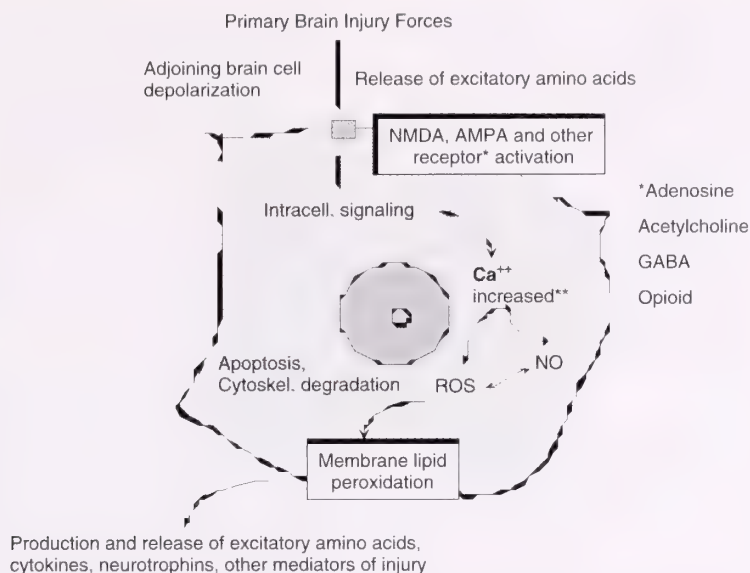


Fig 1. Biochemical responses to primary injury. **Abbreviations:** AMPA, α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; GABA, γ -aminobutyric acid; NMDA, *N*-methyl-D-aspartic acid; NO, nitric oxide; ROS, reactive oxygen species. Reprinted with permission from *Ann Emerg Med*.³⁶ Copyright 2001, American College of Emergency Physicians.

debate about whether high glutamate concentrations following TBI are a product or cause of the initial massive depolarization. Given its pathological excitatory nature, glutamate is a significant contributor to brain cell death and dysfunction in TBI. Alterations in both presynaptic and postsynaptic glutamate receptors are central to this excitotoxicity.¹³ Failure of presynaptic ion pumps and calcium-mediated exocytosis lead to glutamate release. Likewise, traumatized postsynaptic glutamate transmitters have altered function and composition and contribute to TBI excitotoxicity. Glutamate receptors can be broadly categorized as α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors or *N*-methyl-D-aspartic acid (NMDA) receptors. NMDA and AMPA receptors, when activated, normally allow the influx of calcium. Both receptor types have increased calcium influx and neuronal hyperexcitability in the setting of TBI. In experiments on strained neurons, AMPA receptor changes included hyperexcitability, increased ionic currents, and increased intracellular calcium.¹⁴ In animal models, compounds that block NMDA receptors have been associated with improved neurological outcome.

Disruption of calcium homeostasis by glutamate-mediated ion channels, depolarization, or other cellular processes is a key aspect in the progression of secondary injury in TBI. Calcium affects a wide array of cellular functions. Excess calcium leads to disruption of protein phosphorylation, microtubule

construction, protease formation, and a number of other enzyme functions.¹⁵ Sequestration of calcium can also affect the mitochondria. High mitochondrial calcium may result in swelling, depolarization, and a loss of function.¹⁶ This may lead to the initiation of cell death directly via apoptosis or indirectly through loss of oxidative phosphorylation and failure of triphosphate production adenosine (ATP).¹⁷ As ATP provides the energy for the vast majority of cellular functions, including the maintenance of ionic gradients, the inability to generate ATP is tantamount to cell death. Excessive calcium sequestered in the mitochondria also leads to the generation of reactive oxygen species, including the superoxide anion and nitric oxide.¹⁸

NMDA receptor activation is associated with the production of nitric oxide, which can combine with mitochondrially derived superoxide anions to produce the highly reactive nitrating species peroxynitrite. Although nitric oxide is a relatively weak free radical, peroxynitrite is a remarkably powerful one and has been shown to trigger DNA fragmentation, lipid peroxidation, and disruption of amino acids.¹⁹ Oxygen radicals remove a hydrogen atom from unsaturated fatty acids, and this leads to membrane lipid peroxidation. Widespread lipid peroxidation can compromise cellular membranes, resulting in cell lysis and death. Animal studies have demonstrated lipid peroxidation within hours of TBI, and blocking NMDA receptor-mediated nitric oxide

production and peroxynitrite formation has inhibited ischemic infarction in animal models of stroke.²⁰

Oxygen radicals may also be produced by a number of other processes, and their prevalence can be devastating to cells. Depolarization and calcium influx result in an increase in free fatty acids and arachidonic acid. Arachidonic acid is metabolized in the cyclooxygenase pathway, producing lipid hydroperoxides and oxygen radicals. Cerebral hypoperfusion and resultant cellular ischemia may also favor free-radical formation. Catecholamines released in the postinjury period may autoxidate or degrade to produce hydrogen peroxide, which can also participate in oxygen radical reactions. Because of widespread hemorrhage, ample iron is available via the heme complex to catalyze free-radical formation by these processes.

In nonpenetrating TBI, the early postinjury period is characterized by widespread dysfunction of brain tissue, but the structure of neurons, glial cells, and axons is generally preserved.¹² For example, diffuse axonal injury is the result not of large-scale disruption or transection of axonal tracts but of changes on a subcellular level that alter axon function and nerve signaling. Some of these changes are the result of shearing forces that alter neurofilaments and axonal permeability. Axons may undergo axonal disconnection remote from the initial injury. Increased permeability of the axons leads to calcium influx. Increased intracellular calcium is associated with protease activation, microtubule destruction, cytoskeletal breakdown, and axonal disconnection. Axonal injury is a powerful predictor of morbidity and mortality. The loss of axonal function and structure appears to be a major factor in long-term outcome from TBI. Patients with diffuse axonal injury detected by magnetic resonance imaging are more likely to develop persistent coma and are less likely to return to their previous level of functioning.²¹⁻²³

Sites in the injured brain with diffuse axonal injury have been observed to develop regenerative growth cones, but there is no evidence of recovery of neurological outcome.²⁴ Neurogenesis is still limited to the laboratory, but there may be opportunities for pharmacologically mediated recovery of some function.²⁵ White matter injury, as is true for the majority of secondary injuries, is a progressive and delayed process.

The alteration of cellular protein synthesis is a key part of the pathological response and repair following TBI. Cellular injury results in the up-regulation of the genes that code for some proteins and the down-regulation of others. Mediators of injury such as cytokines play a role in modulating cellular protein

production. Posttranslational modification of proteins may produce changes in protein levels that can be detected in serum or cerebrospinal fluid (CSF). This holds some promise in the development of biomarkers to assess the severity of TBI and also to help understand some of the cellular processes that may occur in the injury and postinjury period. Understanding protein regulation and response following TBI may also lead to future interventions to reduce secondary injury.²⁶

A family of enzymes called calpains are another important class of mediators of cellular injury following TBI.²⁷ Calpains are calcium-mediated proteases that under normal cellular conditions have a variety of cellular regulatory functions, including cytoskeletal maintenance. They have been shown to have a significant role in TBI as key enzymes targeting axonal proteins leading to cytoskeletal breakdown and disruption of axonal transport.^{28,29} In animal models, targeted inhibition of these enzymes has shown great promise in limiting secondary injury to axons.³⁰

Regulation of cell cycles and cell death (apoptosis) is a key feature in both the pathogenesis of TBI and protective mechanisms. Although cell death may be the result of a fatal insult to the cell structure or function, it may also be the product of programmed cell death or apoptosis. Caspase is a key enzyme in this pathway and is often a marker of cell death induction. Although it would be easy to assume that cell death would be undesirable, it may also serve a protective role. Likewise, proliferation of cells may be detrimental. The regulation of cell cycles has been an area of intense research, and there are close relationships between TBI and cell cycle regulation. It has been shown that although neurons often die in the setting of TBI, glial cells and astrocytes may proliferate. Inhibitors of apoptotic pathways have in some experiments been demonstrated to improve recovery.³¹

There are numerous other mediators that may play a role in the bimolecular response to TBI. Endogenous opioids, adenosine, catecholamines (dopamine and norepinephrine), and a number of cytokines are among the additional factors in TBI. In most cases, it is unclear if they play a primary role in the neuropathology of TBI or are a reaction to other biological processes. Although interventions have been encouraging in animal models, clinical TBI trials using drugs to reduce the effects of mediators of injury have ubiquitously been unsuccessful. Processes such as apoptosis may have a protective effect in TBI, and in many cases, we cannot be sure that the bimolecular response in question is pathological and not beneficial. Furthermore, brain

tissue is not homogeneous, and what might be pathological to a glial cell may be beneficial to the neuron. There are very significant differences in function and injury response on both macroscopic and cellular levels.

PHYSIOLOGICAL PARAMETERS IN BRAIN INJURY

Given its relative size, the metabolic requirements of the brain are staggering. The brain requires large amounts of oxygen and glucose to remain functional. There are a number of mechanisms by which the human body protects the brain and maintains a constant supply of blood and nutrients. The cranial vault provides protection from the external environment, and the blood-brain barrier (BBB) maintains the integrity of the cerebral milieu, which is distinct from that of the rest of the body. This barrier is created by specialized vascular endothelial cells and a network of tight junctions that prevent the passive diffusion of electrolytes, and large proteins.³² The BBB and autoregulation of CBF are dependent on a number of local, chemical, and endothelial factors and neurotransmitters released from local neurons. The BBB can fail because of a number of insults in the setting of TBI, including ischemic changes to vascular endothelial cells leading to a loss of the integrity of the cerebral environment and failure of autoregulation.

Cerebral autoregulation maintains a relatively constant CBF despite variable cerebral perfusion pressures (CPPs) (See Table 1 for the terminology used in cerebral physiology). CPP is defined as the difference between the mean arterial pressure and ICP. Autoregulation is achieved through tight control of cerebral vascular resistance and is dependent on an intact and functional BBB. CPP below 50 mm Hg leads to brain tissue ischemia and autoregulation failure. When this happens, the brain becomes dependent on the mean arterial pressure for perfusion. Although the brain can compensate for decreased CBF by increasing oxygen extraction, this compensation has finite limits. Therefore, low CPP directly translates into ischemic injury risk.

The brain is very susceptible to ischemic injury, and in the setting of TBI, it is vital for cerebral perfusion and oxygenation to be maintained. Among the recommendations of the Brain Trauma Foundation, the first priority is rapid physiological resuscitation. Hypotension and hypoxia are seen in more than a third of TBI patients and are 2 of the 5 most potent predictors of mortality.³³ Ninety percent

Table 1. *Terminology of Traumatic Brain Injury.*

Abbreviation	Definition
MAP	Mean arterial pressure
ICP	Intracranial pressure
CPP	Cerebral perfusion pressure (mean arterial pressure – intracranial pressure)
CVR	Cardiovascular resistance
CBF	Cerebral blood flow (cerebral perfusion pressure/cardiovascular resistance)

of patients who die from TBI demonstrate ischemic cell changes, and cerebral ischemia may be the most important secondary event affecting outcome following TBI.

CBF is typically low following TBI and is often half of normal in the first 24 hours following injury. In some patients, there is a paradoxical increase in CBF (luxury perfusion) due to a loss of autoregulation. There are a number of factors that contribute to the decrease in CBF following TBI, not only globally but also focally in the brain. CBF is highly dependent on blood pressure, oxygenation, and blood carbon dioxide concentrations. Interventions such as hyperventilation and pharmacological agents can markedly decrease CBF. A mass effect caused by intracranial lesions on cerebral blood vessels, traumatic vasospasm, and decreased metabolism may also play roles. There is a direct correlation between CBF and the Glasgow Coma Scale score in the first 24 hours post-injury, and low CBF is associated with poor outcome and death. By injury type, CBF tends to be the lowest with diffuse injury accompanied by significant edema, with large subdural hematomas, and when the patient has experienced hypotension. TBI patients with normal computed tomography scans and those who have isolated epidural hematoma usually have normal or increased CBF.

The brain is housed in a rigid cavity and has a very limited ability to compensate for hemorrhage, swelling, edema, or mass effects (see Figure 2). Although these pathological processes can often be characterized by diagnostic imaging, clinical examinations are unreliable unless the patient has progressed to brainstem herniation.³⁴ The most accurate way of assessing ICP is to insert a monitor. ICP in the intact cranium is determined by the brain parenchyma tissue pressure, presence of mass lesions, cerebral blood volume, and intracranial CSF volume. Any threshold value for ICP monitoring must take into account that it is the location of intracerebral masses that most directly affects the risk of herniation.

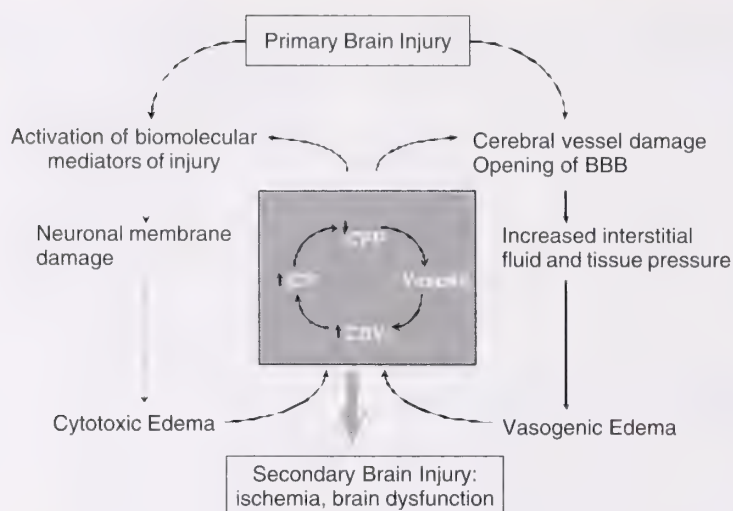


Fig 2. The cycle of primary and secondary injury. **Abbreviations:** BBB, blood-brain barrier; CBV, cerebral blood volume; CCP, cerebral perfusion pressure; ICP, intracranial pressure. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Increased ICPs (intracranial hypertension) can be reduced by the shrinkage of the volume of brain tissue (mannitol or other hypertonic substances), the reduction of CSF (drainage), the reduction of blood (hyperventilation-induced vasoconstriction), or the surgical removal of the pathological process. Alternatively, one may open the skull to allow expansion of the structures (decompressive craniectomy).³⁵ Although ICP monitoring has been a cornerstone of neurosurgical management for decades, it must be always considered in its clinical setting.

Two types of brain edema have been described: interstitial (vasogenic) and intracellular (cytotoxic.) Both types occur following TBI, and both can contribute to secondary injury. Brain edema is usually worst at 24 to 48 hours post-injury. Vasogenic edema can result from failure of the BBB or from reflex vasodilation of vessels. This dilatation can occur the partial pressure of carbon dioxide in the arterial blood increases as a result of ventilatory failure. The increase in cerebral volume that results from vasogenic edema may initially be offset by changes in brain tissue compliance. Compliance may be achieved through the shunting of CSF to the spinal subarachnoid space, an increase in CSF absorption, a decrease in CSF production, or the shunting of venous blood out of the cranium. Although generally linked to ICP, brain edema can often occur without a corresponding increase in ICP.³²

Cytotoxic edema is the result of changes in cellular osmolality with resultant failure of the cell's ability to regulate its ionic gradients. Alterations in cell membrane function and ion transport can

ultimately lead to failure of ATP production and cell death. In areas of the brain with an impaired BBB and ischemia, restoration of blood flow can cause cytotoxic brain injury as well. Neurons and glial cells are particularly susceptible to cytotoxic cell injury. Cytotoxic edema, if widespread, can be associated with increased ICPs, impeded blood flow, and ischemia, yet not independently with permanent neurological impairment.

Hemorrhage in TBI can be from the initial insult or the product of secondary injuries, such as cytotoxic edema, ischemia, or intracranial hypertension. Apart from its effect in increasing ICP, blood is also a rich source of iron, which can catalyze free-radical formation. Larger amounts of bleeding in the epidural or subdural spaces increase intracranial blood volume and pressures. Intraparenchymal and subarachnoid bleeding occurs in the setting of endothelial and BBB disruptions and may exacerbate secondary injury by stimulating the release of excitatory amino acids (glutamate), free radicals, and nitric oxide. Blood within the interstitial space is also proinflammatory.

CONCLUSION

TBI is characterized first by physical forces leading to brain tissue destruction and alteration of function. Although many cells may be rendered dysfunctional during this phase, they are not necessarily irreversibly damaged or disrupted (Table 2). The state of the postinjury cerebral environment and resultant

Table 2. Key Concepts in the Management of Traumatic Brain Injury.

Primary injury is the result of the initial mechanical forces from brain trauma.
Secondary injuries occur over time and are a result of the activation of biomolecular and physiological processes distinct from but synergistic with the primary injury. Excitotoxicity (primarily from glutamate release), alterations in calcium homeostasis, and oxidant damage are key components of secondary injury.
There are no known therapies to inhibit secondary injury processes. The best current approach is rapid resuscitation and avoidance of ischemic injury.
Cerebral autoregulation and the maintenance of the unique cerebral environment are dependent on an intact blood-brain barrier. The function of the blood-brain barrier can be severely altered in the setting of traumatic brain injury.
Vasogenic edema may result from a failure of the blood-brain barrier. Cytotoxic edema is the result of a biomolecular injury and inflammatory cascade that causes loss of ionic gradients, membrane dysfunction, and cellular swelling.

secondary responses to TBI are what ultimately determine the final outcome. If a favorable milieu is created, brain cells and tissues may recover; if it is unfavorable, they may die. The great challenge is to identify what constitutes a damaging environment versus a healing environment and then how to mediate that environment.

An understanding of the biomolecular and physiological responses to TBI is essential to developing future therapies. Much has been learned in this area, but our knowledge is still relatively unsophisticated and incomplete. In the clinical setting, we are currently limited by what we can do on a macroscopic scale. Although there is great promise in the future of secondary interventions, it is unlikely that these will be solitary, standardized therapies; more likely, therapies will need to be tailored to the individual patients, taking into account mechanistic, genetic, cellular, anatomical, and pharmacological differences. With increased understanding of the pathophysiology of brain injury, there is great promise for future therapies; however, to date there have been far more failures than successes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Langlois JA, Rutland-Brown W, Wald MM. The epidemiology and impact of traumatic brain injury: a brief overview. *J Head Trauma Rehabil* 2006; 21: 375–378.
- Jennett B, Teasdale G, Galbraith S, et al. Severe head injury in three countries. *J Neurol Neurosurg Psychiatry* 1977; 40: 291–295.
- Rutland-Brown W, Langlois JA, Thomas KE, Xi YL. Incidence of traumatic brain injury in the United States, 2003. *J Head Trauma Rehabil* 2006; 21: 544–548.
- Foulkes MA, Eisenberg HM, Jane JA, et al. The Traumatic Coma Data Bank: design, methods, and baseline characteristics. *J Neurosurg* 1991; 75: S8–S13.
- Brain Trauma Foundation, American Association of Neurological Surgeons, Joint Section on Neurotrauma and Critical Care. Management and prognosis of severe traumatic brain injury. *J Neurotrauma* 2000; 17: 449–554.
- McLean AJ. Brain injury without head impact? In: Bandak AF, Eppinger RH, Ommaya AK, eds. *Traumatic Brain Injury: Bioscience and Mechanics*. Larchmont, NY: Mary Ann Liebert; 1996; 45–49.
- Teasdale G, Mathew P. Mechanisms of cerebral concussion, contusion and other effects of head injury. In: Youmans JR, ed. *Neurological Surgery*. 4th ed. New York, NY: WB Saunders; 1996; 1533–1546.
- Blumbers PC, Scott G, Manavis J, et al. Staining of amyloid precursor to study axonal damage in mild head injury. *Lancet* 1994; 344: 1055–1056.
- Thibault LE, Gennarelli TA. Brain injury: an analysis of neural and neurovascular trauma in the nonhuman primate. *Annu Proc Assoc Adv Automot Med* 1990; 34: 337–351.
- Willinger R, Taleb L, Kopp C. Modal and temporal analysis of head mathematical models. In: Bandak AF, Eppinger RH, Ommaya AK, eds. *Traumatic Brain Injury: Bioscience and Mechanics*. Larchmont, NY: Mary Ann Liebert; 1996; 265–276.
- Obreja C. Stereotactical phenomena in traumatic brain injury biomechanics: diffuse axonal injury and brain concussion. Available at: http://www.neurostaff.org/intro_en.htm. Accessed January 2008.
- Gentry LR. Imaging of closed head injury. *Radiology* 1994; 191: 1–17.
- van Landeghem FK, Weiss T, Oehmichen M, et al. Decreased expression of glutamate transporters in astrocytes after human traumatic brain injury. *J Neurotrauma* 2006; 23: 1518–1528.
- Goforth PB, Ellis EF, Satin LS. Enhancement of AMPA-mediated current after traumatic injury in cortical neurons. *J Neurosci* 1999; 19: 7367–7374.
- Hayes RL, Dixon CE. Neurochemical changes in mild head injury. *Semin Neurol* 1994; 14: 25–31.
- Verweij BH, Muizelaar JP, Vinas FC, et al. Impaired cerebral mitochondrial function after traumatic brain injury in humans. *J Neurosurg* 2000; 93: 815–820.
- Kim JS, He L, Lemasters JJ. Mitochondrial permeability transition: a common pathway to necrosis and

- apoptosis. *Biochem Biophys Res Commun* 2003; 304: 163–170.
18. Saelens X, Festjens N, Vande Walle L, et al. Toxic proteins released from mitochondria in cell death. *Oncogene* 2004; 23: 2861–2874.
 19. Arundine M, Aarts M, Lau A, et al. Vulnerability of central neurons to secondary insults after in vitro mechanical stretch. *J Neurosci* 2004; 24: 8106–8123.
 20. Aarts M, Liu Y, Liu L, et al. Treatment of ischemic brain damage by perturbing NMDA receptor-PSD-95 protein interactions. *Science* 2002; 298: 846–850.
 21. Maxwell WL, Povlishock JT, Graham DL. A mechanistic analysis of nondisruptive axonal injury: a review. *J Neurotrauma* 1997; 14: 419–440.
 22. Povlishock JT, Kontos HA, Ellis EF. Current thoughts on experimental head injury. In: Becker DP, Gudeman SK, eds. *Textbook of Head Injury*. Philadelphia, PA: WB Saunders; 1989; 451–465.
 23. Povlishock JT. Pathobiology of traumatically induced axonal injury in animals and man. *Ann Emerg Med* 1993; 22: 980–986.
 24. Povlishock JT. Traumatically induced axonal injury: pathogenesis and pathobiological implications. *Brain Pathol* 1992; 1: 1–12.
 25. Hallbergson AF, Gnatenco C, Peterson DA. Neurogenesis and brain injury: managing a renewable resource for repair. *J Clin Invest* 2003; 112: 1128–1133.
 26. Kochanek OM. Berger biomarkers of primary and evolving damage in traumatic and ischemic brain injury: diagnosis, prognosis, probing mechanisms, and therapeutic decision making. *Curr Opin Crit Care* 2008; 14: 135–141.
 27. Huang Y, Wang KK. The calpain family and human disease. *Trends Mol Med* 2001; 7: 355–362.
 28. Carafoli E, Molinari M. Calpain: a protease in search of a function? *Biochem Biophys Res Commun* 1998; 247: 193–203.
 29. Ai J, Liu E, Wang J, et al. Calpain inhibitor MDL-28170 reduces the functional and structural deterioration of corpus callosum following fluid percussion injury. *J Neurotrauma* 2007; 24: 960–978.
 30. Saatman KE, Murai H, Bartus RT, et al. Calpain inhibitor AK295 attenuates motor and cognitive deficits following experimental brain injury in the rat. *Proc Natl Acad Sci U S A* 1996; 93: 3428–3433.
 31. Di Giovanni S, Movsesyan V, Ahmed F, et al. Cell cycle inhibition provides neuroprotection and reduces glial proliferation and scar formation after traumatic brain injury. *Proc Natl Acad Sci U S A* 2005; 102: 8333–8338.
 32. Betz AL, Crockard A. Brain edema and the blood brain barrier. In: Crockard A, Hayward R, Hoff JT, eds. *Neurosurgery: The Scientific Basis of Clinical Practice*. 2nd ed. Oxford, England: Blackwell Scientific; 1992; 353–372.
 33. Chesnut RM, Marshall LF, Klauber MR, et al. The role of secondary brain injury in determining outcome from severe head injury. *J Trauma* 1993; 34: 216–222.
 34. Lundberg N. Continuous recording and control of ventricular fluid pressure in neurosurgical practice. *Acta Psychiatry Neurol Scand* 1960; 36(suppl 149): 1–193.
 35. Vincent JL, Berre J. Primer on medical management of severe brain injury. *Crit Care Med* 2005; 33: 2157.
 36. Zink BJ. Traumatic brain injury outcome: concepts for emergency care. *Ann Emerg Med* 2001; 37: 318–332.

Traumatic Brain Injury in the United States: An Epidemiologic Overview

Carl R. Summers, PhD, Brian Ivins, MS, and Karen A. Schwab, PhD

Defense and Veterans Brain Injury Center, Washington, DC

ABSTRACT

A basic description of severity and frequency is needed for planning healthcare delivery for any disease process. In the case of traumatic brain injury, severity is typically categorized into mild, moderate, and severe with information from a combination of clinical observation and self-report methodologies. Recent US civilian epidemiological findings measuring the frequency of mortality and morbidity of traumatic brain injury are presented, including demographic and etiological breakdowns of the data. Falls, motor vehicle accidents, and being struck by objects are the major etiologies of traumatic brain injury. US civilian and Army hospitalization trends are discussed and compared. Features of traumatic brain injuries from Operation Iraqi Freedom and Operation Enduring Freedom are discussed. *Mt Sinai J Med* 76:105–110, 2009. © 2009 Mount Sinai School of Medicine

Key Words: epidemiology, head injury, traumatic brain injury.

A Centers for Disease Control and Prevention analysis of hospital, emergency department (ED), and vital statistics databases estimated that about 1.4 million people presented for medical care for a traumatic brain injury (TBI) each year in the United States from 1995 through 2001.¹ The analysis also found that approximately 50,000 (3.6%) of them died from their injuries, 235,000 (17%) were hospitalized, and 1.1 million (80%) were treated and released from the ED.¹ A separate study using data about people hospitalized with TBI estimated that at the beginning

of 2005, 3.17 million people in the United States (1.1% of the total population) were living with long-term disability that resulted from TBI.² The number of people with TBI who present in outpatient settings other than EDs, such as physicians' offices, is currently unknown as is the number of those with TBI who do not seek medical attention. However, the proportions of those with TBI who seek medical attention outside the ED or who do not seek any medical attention may be sizable. A study of data obtained from the 1991 National Health Interview Survey estimated that 25% of individuals with a self-reported TBI that resulted in loss of consciousness did not seek medical attention and that 14% were evaluated in clinics or offices.³

These findings reveal 2 important features of TBI. One is that it is a common injury. The other is that TBI outcomes vary greatly. They also indicate that many people who have milder forms of TBI are not identified in current databases. Studying the epidemiology of TBI is challenging for a number of reasons, including limited data sources and different methods for classifying TBI severity. However, despite these challenges, enough research has been performed about the epidemiology of TBI to identify the most common mechanisms of injury and to characterize TBI risk in various segments of the population. This article summarizes findings from epidemiological studies of TBI in the United States that provide national-level data and briefly discusses some methodological considerations that can help readers of TBI epidemiology papers understand why particular methodologies were used. A brief review of recent research findings about TBI in the US military is included.

Address Correspondence to:

Carl R. Summers

Defense and Veterans Brain
Injury Center

Washington, DC

Email: Carl.Summers@amedd.army.mil

METHODS

We searched PubMed and Medline for articles published in the last 10 years that report national-level data about the epidemiology of TBI in the

United States. Search terms included "traumatic brain injury" paired with each of the following:

- Epidemiology incidence trends (124 citations found).
- Severity classification (487 citations found).
- Symptom prevalence (78 citations found).
- Military (566 citations found).

The abstracts of publications that appeared to be relevant were then reviewed. A total of 30 publications were found and reviewed. In addition, several seminal articles from outside this timeframe were included because they established current definitions used for TBI severity^{4,5} or reported seminal evidence on specific topics.³

INCIDENCE OF TRAUMATIC BRAIN INJURY

Overall, falls were the leading cause of TBI in the United States each year from 1995 through 2001.¹ Falls accounted for an annual average of 398,000 (28%) TBI cases during that period. Automobile crashes were the second leading cause of TBI, accounting for an average of 280,000 (20%) cases per year.¹ Being struck by or against something was the third leading cause, accounting for 19% of TBI cases annually.¹ Another common cause of TBI is assault, which accounted for 156,000 TBI cases per year.¹ A number of TBI cases do not have a known cause. These amounted to an average of 9% of all TBI cases per year from 1995 through 2001.¹

The incidence of TBI among males is higher than it is among females. Males account for 60% of TBI cases that occur each year. Young children, older adolescents, and the elderly have the highest TBI rates.¹ Fall-related TBI among the elderly is particularly problematic. TBI accounts for 50.3% of unintentional fall deaths and 8.0% of nonfatal fall-related hospitalizations among adults over 65 in the United States.⁶

CIVILIAN TRENDS

The published literature has shown that the incidence of TBI mortality and hospitalization rates in the United States decreased in the 1980s and 1990s. TBI-related mortality in the United States, for example, declined 22.6% (from 24.6/100,000 to 19.3/100,000) from 1979 to 1992; there was an 11.4% decline (from 21.9/100,000 to 19.4/100,000) from 1989 to 1998.⁷ Probable causes cited in the literature for

these decreases include injury prevention efforts, compulsory safety laws (eg, seatbelt, helmet, and drunk driving laws), engineered solutions (eg, air bags), and improved treatment for alcohol and drug problems.⁸

An analysis of data from the National Hospital Discharge Survey, Thurman and Guerrero⁹ found that the incidence of hospitalization associated with civilian TBI decreased 51% from 1980 to 1995 (from 199/100,000 to 98/100,000). However, most of this was due to a decrease in the incidence of hospitalizations for mild TBI. They found that hospitalizations for mild TBI decreased 61% (from 130/100,000 to 51/100,000) during that period, whereas the incidence of hospitalization associated with moderate TBI decreased 19% (from 26/100,000 to 21/100,000). The incidence of hospitalization for severe TBI increased 90% during that period (from 10/100,000 to 19/100,000).

Possible reasons for changes in TBI hospitalization rates include the following:

- Changes in hospital admission practices for mild TBI due in part to changes in health insurance coverage.⁹
- Increased survivability after severe TBI: With improved availability and quality of TBI treatment, an increasing proportion of potentially fatal TBI incidents are survivable. This is a reasonable explanation for the downward trend of TBI mortality rates with a concurrent increase in severe TBI hospitalization rates.
- Improved safety and prevention programs: Part of the decrease in both mortality and morbidity in the 1980s and 1990s may be due to injury prevention measures, especially those associated with motor vehicles⁹ and motorcycle helmet laws.

However, more recent research suggests that the incidence of death due to TBI is no longer decreasing at those rates. Rutland-Brown and colleagues¹⁰ found that the TBI mortality rate was 18.1/100,000 in 1998 and 17.9/100,000 in 2003. Data from the latest report from the Center for Disease Control also indicate that the rate of decrease in TBI hospitalizations has also leveled off.¹⁰

MILITARY TRENDS

The incidence of TBI-related hospitalization rates in the active-duty US Army decreased during the 1990s. The overall incidence decreased 75% (from 247.7/100,000 to 62.3/100,000) from 1990 to 1999.¹¹ The hospitalization rate for mild TBI in the Army

Table 1. *Glasgow Coma Scale Scoring Rubric.*

		Best Response Score					
Test	1	2	3	4	5	6	
Adults							
Eyes	Does not open eyes	Opens eyes in response to painful stimuli	Opens eyes in response to voice	Opens eyes spontaneously	N A	N A	
Verbal	Makes no sounds	Incomprehensible sounds	Utters inappropriate words	Confused, disoriented	Oriented, converses normally	N A	
Motor	Makes no movements	Extension to painful stimuli	Abnormal flexion to painful stimuli	Flexion withdrawal to painful stimuli	Localizes painful stimuli	Obeys commands	
Children							
Eyes	No eye opening	Eyes opening to pain	Eyes opening to speech	Eyes opening spontaneously	N A	N A	
Verbal	No verbal response	Inconsolable, agitated	Inconsistently inconsolable, moaning	Cries but consolable; inappropriate interactions	Smiles, orients to sounds, follows objects, interacts	N A	
Motor	No motor response	Extension to pain (decelerate response)	Abnormal flexion to pain for an infant	Infant withdraws from pain	Infant withdraws from touch	Infant moves spontaneously or purposefully	

NOTE: This table was adapted from Teasdale and Jennett¹ and Camney *et al.*²⁵**Abbreviation:** N A, not applicable.

decreased 78.6% (from 159/100,000 to 34.0/100,000), whereas the rate for moderate TBI decreased 57.1% (from 14.3/100,000 to 6.1/100,000), and the rate for severe TBI decreased 53.7% (from 22.8/100,000 to 10.6/100,000). A likely reason for the decrease in the Army's TBI-related hospitalization rates is a change in admission practices for mild TBI similar to those that occurred in civilian hospitals. Another possible reason is an increased emphasis on injury prevention during the 1990s.¹² Studies from the 1990s indicated that accident-related injuries in the Army decreased to record lows.¹³

A new TBI-related concern for both the military and civilian medical establishment is the influx of TBI cases due to the military involvement in the Iraq [Operation Iraqi Freedom (OIF)] and Afghanistan [Operation Enduring Freedom (OEF)] theaters of operation. These cases, including members of the National Guard and Reserve, are handled initially through the Department of Defense and Department of Veterans Affairs; however, some war-related TBI patients are likely to be seen by civilian providers. Civilian providers in some localities might see nontrivial increases in TBI cases from Reserve and National Guard combat units returning from OEF/OIF as well as OEF/OIF combat veterans who have been released from active duty and choose to be treated by private healthcare providers.

As a group, OEF/OIF TBI patients are likely to present with 2 common characteristics. The first characteristic is that the etiology of their injuries is likely to involve a blast (explosion). Gondusky and Reiter¹⁴ reported that in their sample of a returning combat battalion, 97% of the 188 injuries sustained by 120 Marines were from improvised explosive devices (65%) or mines (32%). Using a survey of 2525 soldiers 3 to 4 months after deployment, Hoge *et al.*¹⁵ reported that 79% of injuries reported in their survey of returned OEF/OIF veterans were from blasts. Although there is evidence that a large percentage of service members who have experienced combat-related TBI have done so because of blast wounds (frequently in combination with other causes of injury such as vehicle crashes or falls), it is not known if the TBI blast sequelae are materially different from TBI resulting only from noncombat causes such as falls and vehicle crashes. In many cases, multiple injury mechanisms are involved, and it is not possible to determine the exact cause of the TBI.

The second characteristic is that service members who are injured in combat with a TBI also have a high occurrence of posttraumatic stress disorder (PTSD) and depression. Hoge and colleagues,¹⁵ using a sample of US Army soldiers who served in Iraq, found that 43.9% of service members that had

experienced loss of consciousness (LOC) screened positive for PTSD and 24.8% screened positive for depression. Civilians with TBI have similar psychiatric comorbidities. Silver and colleagues¹⁶ found that 11.1% of adults who screened positive for self-reported TBI also had a history of depression, but the prevalence among OIF/OEF veterans may be higher.

METHODOLOGICAL CONSIDERATIONS THAT INFLUENCE RESULTS OF TRAUMATIC BRAIN INJURY EPIDEMIOLOGY STUDIES

A variety of methods have been used to perform TBI epidemiology studies. Perhaps the most important methodological consideration faced in any type of TBI research is how to classify severity. Although TBIs are generally classified as being mild, moderate, or severe, different methods can be used to determine severity. The severity classification method that is used depends largely on the source of the data for the study. This section of the article provides a brief overview of different ways in which TBI severity can be classified and when they are likely to be used.

One commonly used method is to base severity on the Glasgow Coma Scale (GCS) score (Table 1).^{4,25} The GCS score is a measure of an individual's level of consciousness. It assesses 3 domains: eye opening, motor response, and verbal response. Patients can receive a score of 1 to 4 for eye opening, 1 to 6 for motor response, and 1 to 5 for verbal response. The scores from each domain are summed to obtain the GCS score, which ranges from 3 to 15. TBI patients with a GCS score of 3 to 8 are considered to have severe TBI, those with a score of 9 to 12 are considered to have moderate TBI, and those with a score of 13 to 15 are considered to have mild TBI.

Another method is to base severity on the duration of LOC and/or posttraumatic amnesia (PTA). TBIs that result in LOC lasting 30 minutes or less and/or PTA lasting less than 24 hours are defined as mild by the American Congress of Rehabilitation, the Centers for Disease Control and Prevention, and the World Health Organization.^{5,17,18} TBIs that result in LOC or PTA of longer duration are considered moderate or severe.

TBI severity can also be determined with Abbreviated Injury Scale (AIS) scores.¹⁹ The AIS classifies injury severity into 1 of 6 categories: 1, mild; 2, moderate; 3, serious; 4, severe; 5, critical; and 6, maximal (injuries that cannot be treated). Researchers who have used AIS scores to determine TBI severity have collapsed these into the mild, moderate, and

Table 2. *Key Concepts.*

Approximately 1.4 million Americans are treated for traumatic brain injury annually.

Traumatic brain injury is typically categorized as mild, moderate, or severe with clinical observation (which may include imaging) and self-report methodologies.

Traumatic brain injury is a common injury that occurs most frequently in young males. Its most common etiologies are falls, motor vehicle accidents, and being struck by or against an object.

The rate of traumatic brain injury mortality and hospitalizations generally fell in the 1980s and 1990s but has leveled off since. Most of the drop in hospitalizations is in the mild traumatic brain injury range and is believed to be due, at least in part, to changes in hospital admission practices.

A developing concern for traumatic brain injury treatment is a nontrivial number of wounded Iraq/Afghanistan war veterans returning with traumatic brain injury. Members of this group are likely to have been wounded by a blast and have an elevated chance of experiencing posttraumatic stress disorder and/or depression.

severe designations that are generally used to classify TBI. When this is done, TBIs with AIS scores of 1 or 2 are considered mild, those with an AIS score of 3 are considered moderate, and those with AIS scores of 4 to 6 are considered severe.

The method used to determine TBI severity in a study is usually dictated by the source of data being used. When a researcher has access to full medical records, GCS can often be used to determine severity. When large administrative databases are used, such as the National Hospital Discharge Survey, severity can be determined with AIS scores. When data are obtained from self-report questionnaires, severity can be determined from the reported duration of LOC or PTA.

LIMITATIONS

Epidemiological research in TBI is fraught with well-documented limitations.^{20–24} Definitional inconsistency, limited data, practical data collection constraints, limited resources, nontrivial numbers of injured people who do not seek treatment (resulting in an undercount), and cases of TBI being overlooked or not properly documented are the major limitations. Thus, any inferences made from these data should be made with these considerations in mind.

CONCLUSIONS

Overall TBI mortality rates in the US civilian population declined in the 1980 and 1990s but have since stabilized. TBI hospitalization rates in the US civilian population followed the same basic trajectory as mortality for the same period; however, analysis by severity shows that severe TBI incidence rates rose, moderate TBI rates were level, and mild TBI hospitalizations dropped dramatically. These trends are consistent with 3 possible scenarios: changes in admission practices for mild TBI,

improvements in the delivery and care of treatment to very severe patients (who previously would have died), and improvements in safety and prevention (Table 2).

Overall US Army TBI hospitalizations dropped during the 1990s; however, unlike their civilian counterparts, a substantial decrease in TBI hospitalization was reported for all severity levels. Ivins *et al.*¹¹ suggested that possible reasons for the decrease were changes in hospital admission policies for mild TBI and injury prevention programs. However, it is likely that TBI hospitalizations in the Army have increased because of the wars in Iraq and Afghanistan.

The influx of troops with TBI returning from the Iraqi and Afghanistan conflicts has created a new concern for the US medical system. A large percentage of returning service members with TBI have been exposed to blasts, often in combination with other causes of injury. It is not known if there are specific sequelae caused by TBI blast exposure or if they are substantially similar to TBI sequelae resulting from noncombat causes such as falls and vehicle accidents. Additional considerations for this population include elevated rates of PTSD and depression.

ACKNOWLEDGMENT

The views expressed in the article are those of the authors and do not reflect the official policy or position of the Department of Defense or the US Government.

DISCLOSURES

Potential conflict of interest: None.

REFERENCES

1. National Center for Injury Prevention and Control. *Traumatic Brain Injury in the United States:*

- Emergency Department Visits, Hospitalizations, and Deaths*. Atlanta, GA: Centers for Disease Control and Prevention; 2006.
2. Zaloshnja E, Miller T, Janglois JA, Selassie AW. Prevalence of long term disability from traumatic brain injury in the civilian population of the United States. *J Head Trauma Rehabil* 2008; 23: 394–400.
 3. Sosin DM, Sniezek JE, Thurman DJ. Incidence of mild and moderate brain injury in the United States, 1991. *Brain Inj* 1996; 10: 47–54.
 4. Teasdale G, Jennett B. Assessment of coma and impaired consciousness, a practical scale. *Lancet* 1974; 2: 81–84.
 5. Kay T, Harrington DE, Adams R, et al. Definition of mild traumatic brain injury. *J Head Trauma Rehabil* 1993; 8: 86–87.
 6. Thomas KE, Stevens JA, Sarmiento, K, Wald MM. Fall-related traumatic brain injury deaths and hospitalization among older adults—United States, 2005. *J Saf Res* 2009; 39: 269–272.
 7. Adekoya N, Thurman DT, White DJ, Webb KW. Surveillance for traumatic brain injury deaths: United States, 1989–1998. *MMWR Surveill Summ* 2002; 51: 1–14.
 8. Napaolitano E, Radecki J, Elovic EP. Brain injury medicine: principles and practice. In: Zasler ND, Katz DI, Zafonte RD, eds. *Brain Injury Medicine*. New York, NY: Demos Medical Publishing; 2007; 57–79.
 9. Thurman D, Guerrero J. Trends in hospitalization associated with traumatic brain injury. *JAMA* 1999; 282: 954–957.
 10. Rutland-Brown W, Langlois JA, Thomas KE, Yongli LX. Incidence of traumatic brain injury in the United States, 2003. *J Head Trauma Rehabil* 2006; 21: 544–548.
 11. Ivins BJ, Schwab KA, Baker G, Warden DL, Moss NG. Hospital admissions associated with traumatic brain injury in the US Army during peacetime: 1990 trends. *Neuroepidemiology* 2006; 27: 154–163.
 12. Brundage JF. Military preventive medicine and medical surveillance in the Post-Cold War era. *Mil Med* 1998; 163: 272–277.
 13. *Army Achieves Safest Year on Record*. Washington, DC: US Army Public Affairs; 1997.
 14. Gondusky JS, Reiter MP. Protecting military convoys in Iraq: an examination of battle injuries sustained by a mechanized battalion during Operation Iraqi Freedom II. *Mil Med* 2005; 170: 546–549.
 15. Hoge CW, McGurk D, Thomas JL, et al. Mild traumatic brain injury in U.S. soldiers returning from Iraq. *N Engl J Med* 2008; 358: 453–463.
 16. Silver JM, Kramer, R, Greenwald S, Weissman M. The association between head injuries and psychiatric disorders: findings from the New Haven NIMH Epidemiologic Catchment Area Study. *Brain Inj* 2001; 15: 935–945.
 17. National Center for Injury Prevention and Control. *Report to Congress on Mild Traumatic Brain Injury in the United States: Steps To Prevent a Serious Public Health Problem*. Atlanta, GA: Centers for Disease Control and Prevention; 2006.
 18. Carrol LJ, Cassidy JD, Holm L, et al. Methodological issues and research recommendations for mild traumatic brain injury: the WHO Collaborating Center Task Force on Mild Traumatic Brain Injury. *J Rehabil Med* 2004; 43(O:suppl): 113–125.
 19. Association for the Advancement of Automotive Medicine. *Abbreviated Injury Scale, 2005*. Des Plaines, IL: Association for the Advancement of Automotive Medicine; 2005.
 20. Saatman KE, Duhaime A-C, Bullock R, et al. Classification of traumatic brain injury for targeted therapies. *J Neurotrauma* 2008; 25: 719–738.
 21. Thurman DJ, Coronado V, Selassie A. The epidemiology of TBI: implications for public health. In: Zasler ND, Katz DI, Zafonte RD, eds. *Brain Injury Medicine*. New York, NY: Demos Medical Publishing; 2007.
 22. Setnik L, Bazarian JJ. The characteristics of patients who do not seek medical treatment for traumatic brain injury. *Brain Inj* 2007; 21: 1–9.
 23. Moss NG, Wade GT. Admission after head injury: how many occur and how many are recorded. *Injury* 1996; 27: 159–161.
 24. Powell JM, Ferraro JV, Dikmen SS, et al. Accuracy of mild traumatic brain injury diagnosis. *Arch Phys Med Rehabil* 2008; 89: 1550–1555.
 25. Cane N, Badiatia N, Crocco TJ, et al. *Guidelines for Prehospital Management of Traumatic Brain Injury*. 2nd ed. New York, NY: Brain Trauma Foundation; 2007.

Blast-Related Mild Traumatic Brain Injury: Mechanisms of Injury and Impact on Clinical Care

Gregory A. Elder, MD and Adrian Cristian, MD

Rehabilitation Medicine Service, James J. Peters Department of Veterans Affairs Medical Center, Bronx, NY

ABSTRACT

Mild traumatic brain injury has been called the signature injury of the wars in Iraq and Afghanistan. In both theaters of operation, traumatic brain injury has been a significant cause of mortality and morbidity, with blast-related injury the most common cause. Improvised explosive devices have been the major cause of blast injuries. It is estimated that 10% to 20% of veterans returning from these operations have suffered a traumatic brain injury, and there is concern that blast-related injury may produce adverse long-term health affects and affect the resilience and in-theater performance of troops. Blast-related injury occurs through several mechanisms related to the nature of the blast overpressure wave itself as well as secondary and tertiary injuries. Animal studies clearly show that blast overpressure waves are transmitted to the brain and can cause changes that neuropathologically are most similar to diffuse axonal injury. One striking feature of the mild traumatic brain injury cases being seen in veterans of the wars in Iraq and Afghanistan is the high association of mild traumatic brain injury with posttraumatic stress disorder. The overlap in symptoms between the disorders has made distinguishing them clinically challenging. The high rates of mild traumatic brain injury and posttraumatic stress disorder in the current operations are of significant concern for the long-term health of US veterans with associated economic implications. ***Mt Sinai J Med* 76:111–118, 2009.** © 2009 Mount Sinai School of Medicine

Key Words: animal models, blast overpressure wave, blast-related brain injury, improvised explosive device, mild traumatic brain injury, Operation Enduring Freedom, Operation Iraqi Freedom, pathophysiology, postconcussion syndrome, posttraumatic stress disorder, shell shock.

More than 1.6 million soldiers have served in the wars in Iraq [Operation Iraqi Freedom (OIF)] and Afghanistan [Operation Enduring Freedom (OEF)].¹ In both theaters of operation, traumatic brain injury (TBI) has been a major cause of mortality and morbidity, with blast-related injury the most common cause.² Although initial reports focused on soldiers with moderate-to-severe TBIs that had been initially identified in theater,³ it soon became clear that many OIF/OEF veterans were presenting to Veterans Administration (VA) hospitals with symptoms consistent with mild traumatic brain injuries (MTBIs) that were never recognized prior to discharge. These cases are only now being recognized as related to single or often multiple blast exposures. Although the extent of the problem is not fully known, it is without doubt substantial, and estimates are that 10% to 20% of returning OIF/OEF veterans have suffered a TBI.^{1,4} There is a growing concern that blast-related injury may produce long-term affects and affect the resilience and in-theater performance of active duty troops. Indeed, TBI has been called the signature injury of the wars in Iraq and Afghanistan.⁵

To deal with the problem of missed cases, the Department of Defense has implemented population-based screening procedures for MTBI, and troops in the field who may have experienced a blast-related TBI are evaluated by medics using a structured assessment tool, the Military Acute Concussion Evaluation (MACE).⁶ In addition, the VA has mandated TBI screening for all OIF/OEF veterans who present to VA hospitals for any reason, and the VA has established a polytrauma/TBI system

Address Correspondence to:

Gregory A. Elder

Rehabilitation Medicine Service
James J. Peters Department of
Veterans Affairs Medical Center
Bronx, NY

Email: gregory.elder@va.gov

of care nationwide to deal with identified cases.⁷ How a blast affects the brain has also become a question of some urgency, and last summer, the US Congress committed \$150 million administered through the Department of Defense for TBI research during fiscal year 2008, with additional funds being committed for 2009.⁸

METHODOLOGY

To supplement the authors' own knowledge, a PubMed search was performed with the search terms "blast" and "traumatic brain injury." Much information related to blast-related injury in the OIF/OEF campaigns is also available in a recent Rand Corporation report¹ and on the Web site of the Defense and Veterans Brain Injury Center.⁶ A major limitation of the current literature is the paucity of information on the pathophysiology of MTBI in general and blast-related injury in particular, which may differ in significant ways from the mechanisms associated with closed impact head injury typical of civilian TBI.

GENERAL MECHANISMS OF BLAST-RELATED INJURY

Blast injuries can result from a variety of exposures, including mortar shells, rocket-propelled grenades, and other forms of improvised explosive devices (IEDs). However, IEDs have proven to be the major cause of blast injuries in Iraq and Afghanistan and have been estimated to be responsible for approximately 40% of coalition deaths in Iraq and to account for a roughly similar percentage of TBIs in returning soldiers.² Although diverse in design, IEDs typically consist of an explosive charge made from commercially manufactured or homemade explosives coupled to a detonator. IEDs may also incorporate shrapnel-generating materials such as nails or other metal particles.

A high-explosive detonation causes the nearly instantaneous conversion of a solid or liquid explosive material into a gas.^{9,10} The gases that are produced expand rapidly, forming a high-pressure wave that travels at supersonic speeds; this is known as the blast overpressure wave. Because of the rapid expansion of gases, after the initial overpressure wave, pressure drops; this creates a relative vacuum or blast underpressure wave that results in a momentary reversal of air flow or reversed blast wind, which is in turn followed by a second

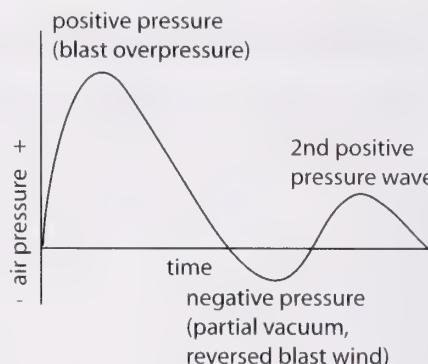


Fig 1. Components of the blast pressure wave. As gases expand rapidly, a high-pressure wave traveling at supersonic speeds is generated (the blast overpressure wave). As the pressure drops, a relative vacuum is created that momentarily leads to a reversal of air flow or reversed blast wind. A second lower intensity positive pressure wave follows before atmospheric pressure returns to normal. See Taber *et al.*¹⁰ for additional discussion.

positive pressure wave before atmospheric pressure returns to normal (Figure 1).

Blast-related injury can occur through several mechanisms.⁹ *Primary blast injury* refers to effects from the blast overpressure wave itself. The primary blast wave generates extreme pressure changes that can cause stress and shear injuries. Organs are affected differentially, with those having air-fluid interfaces generally being the most vulnerable.⁹ For example, rupture of the tympanic membranes is particularly common after blast injury, and the lungs and gastrointestinal tract are also regarded as highly sensitive.⁹ In addition, the blast wave that follows the overpressure wave can propel objects such as shrapnel contained within the IED with considerable force, causing what is called *secondary injury*. The blast wave may also cause the individual to be knocked down or blown into solid objects, resulting in what is called *tertiary injury*. Finally, the heat produced by IED explosions may cause burns, or injuries may result from the inhalation of noxious gases; this last mechanism is called *quaternary injury*.

Studies in Experimental Animals

Experimental animals have been exposed to various forms of blasts ranging from direct exposure to live explosives to controlled blast waves delivered in a tube by a pressure generator. Although live explosives may best model exposure in the field, this approach suffers from the lack of experimental control as well as the problem of separating primary injury from secondary injury. Pressure generators allow effects of the primary blast wave itself to be studied and offer more experimental control. By the

application of appropriate shielding, it is also possible to isolate the effects of blast injury to the body from injury to the brain.

Studies in rats have established that the primary blast wave is transmitted through the skull to the brain.¹¹ In pigs, a transient flattening of the electroencephalogram with apnea has been observed immediately after blast exposure; this effect could be indicative of a blast effect on the brainstem.¹² Other studies have described pulmonary pathology following blast injury in rats¹³ and rabbits.¹⁴ Cernak *et al.*^{14–16} described central nervous system (CNS) injury after whole body or primary lung blast injury in rats, including neuronal, glial, and myelin abnormalities, along with evidence for the activation of inducible nitric oxide synthase (iNOS) and evidence of oxidative stress in the brain. Functionally, the animals showed impairments in a previously learned active avoidance task.^{15,16} Both types of blast exposure resulted in ultrastructural changes, including expanded perineuronal spaces, the appearance of cytoplasmic vacuoles, myelin deformation, and axoplasmic shrinkage. These effects were thought to be most reminiscent of the neuronal abnormalities seen in diffuse axonal injury (DAI). Other studies in rats have supported a likely role for disrupted axonal transport, finding transient alterations in the distribution of neurofilament proteins in cortical regions following blast injury to the head.¹⁷

One limitation of most animal studies to date is that they have used relatively powerful blast exposures (ie, 150–400 kPa) delivered to anesthetized animals; the levels have been high enough to induce significant pulmonary pathology.^{15,16} More recently, Moochhala *et al.*¹⁸ studied the effects of lower blast pressures (2.8–20 kPa). Although the blast-exposed animals showed no deficits in a passive avoidance learning task, the 20-kPa-exposed animals showed impaired motor functions and histologically scattered hyperchromatic and apoptotic neurons in the cerebral cortex 24 hours after exposure. In these studies, an iNOS-related mechanism for the neuronal damage was suggested by the fact that aminoguanidine, a known iNOS inhibitor, was found to be neuroprotective against the injuries. Collectively, these studies seem to clearly indicate that the primary blast wave has CNS effects and that these effects may be apparent at even relatively modest blast pressures.

Mechanisms of Blast-Related Central Nervous System Injury in Humans

The human brain can clearly be affected by the secondary or tertiary injury associated with blast

exposure. Close exposure to a high-pressure blast wave can cause moderate-to-severe TBI, although these injuries are typically a mixture of secondary and tertiary injuries as well, and this makes the contribution of the primary blast wave difficult to ascertain. What is less clear is the degree to which long-term effects can be caused by a primary blast wave sufficient to produce MTBI and whether multiple exposures to this type of blast, which is common among troops in Iraq, can lead to significant long-term injury.

The best understood pathophysiological mechanisms associated with TBI are bleeding, direct tissue damage, and DAI.¹⁹ DAI results when angular forces cause shearing or stretching of axons, which can lead to impaired axonal transport resulting pathologically in the appearance of focal axonal swellings. DAI is common after closed head injuries and most commonly affects tracts at gray/white matter junctions, particularly in the frontal and temporal regions. Contusions occur as the result of coup/contrecoup injury, most affecting the frontotemporal regions and occipital lobes.

All of these mechanisms can be seen with moderate-to-severe blast-related TBI, although as noted previously, these cases are typically a mixture of primary, secondary, and tertiary injuries. Blast waves can also cause MTBI without significant secondary or tertiary injury, the basis of which is presumed to be a pathophysiological mechanism most similar to DAI. However, in truth, the mechanisms underlying these effects are not well understood. We have little neuropathological data on MTBI in general and none related to blast-induced MTBI, with conventional neuroimaging being typically normal in blast-related MTBI cases.

Recently, a peripheral mechanism related to blast-induced vascular changes has also been proposed.²⁰ Specifically, it has been suggested that a high-pressure blast wave hitting the body may induce oscillating high-pressure waves that can be transmitted through the systemic circulation to the vasculature of the brain. The theory proposes that pressure waves may damage structures close to cerebral vessels, including axonal fibers and cells. Although at present there is no direct experimental support for the theory, the hypothesis itself is certainly novel and differs radically from the mechanisms presumed to underlie the acceleration/deceleration injuries found in closed impact injuries typical of civilian head trauma. If correct, it also has practical implications for the protection of troops in the field in that even blast-resistant helmets would not protect against brain injury and instead new types of body armor would be needed that could absorb or deflect

blast wave energies in order to protect against CNS injury.

DIAGNOSIS OF BLAST-RELATED MILD TRAUMATIC BRAIN INJURY

As in civilian head trauma, the diagnosis of blast-related MTBI is largely based on establishing an accurate history of the specific events that led to the brain injury. It is important to determine if there was direct trauma to the brain (ie, falls, motor vehicle accidents, or bullets) or indirect trauma (ie, close exposure to blasts). Brain injury from blasts is often difficult to determine. Soldiers in these wars are in constant exposure to blasts, some close and some at a distance. In addition, it is often difficult for a veteran to recall the events surrounding a blast that may have occurred several months in the past.

It is important to obtain as many details about specific events as possible (see Table 1). Good detective work is often necessary. Obtaining medical records or talking to key people around the event can be challenging; however, it can yield useful information. Troops in the field who have experienced a blast injury are evaluated for TBI by medics using a structured assessment tool known as MACE.⁶

MACE consists of 13 items (8 history items and 5 examination items). In the history assessment, the clinician records details of the incident. Items such as a description of the event, cause of the injury (ie, explosion, blast, fall, motor vehicle accident, or gunshot wound), wearing of a helmet, amnesia surrounding the event, loss of consciousness (LOC), and symptoms are recorded. In the examination, the clinician evaluates 5 domains: (1) orientation; (2) immediate memory;

(3) neurological screening of vision, speech, and motor function; (4) concentration, and (5) delayed recall. The total maximum score for the examination is 30. Scores below 25 may be indicative of cognitive impairment.⁶ The utility of this instrument for predicting subsequent functional impairment has not been established; however, it holds the potential for providing the most objective history regarding the initial injury.

It is clearly important to establish the presence of current symptoms such as headaches, dizziness, impaired balance, memory loss, inability to focus, difficulty making decisions, mental fatigue, irritability, and problems sleeping soon after the event and their persistence to the present day. In addition, the impact of these symptoms on the current functional level and relationships should be ascertained: For example, are they having difficulties at work keeping up with their job requirements or getting along with supervisors and coworkers? If in school, are they receiving low grades in course work? The impact of these symptoms on key social relationships with spouses, children, and friends should also be determined.

Regrettably, the physical examination is often limited in making the diagnosis of MTBI in blast-injured veterans, especially if they are removed from the immediate event. Nevertheless, a thorough neurological and cognitive examination should be performed that assesses orientation, concentration, immediate and delayed recall, motor strength, balance, and gait. An assessment for tympanic membrane rupture is also important because this has been described as a marker of concussive blast injury in soldiers.²¹ Standard neuroimaging such as computed tomography scanning and magnetic resonance imaging are often not helpful but should be considered on a case-by-case basis.

Distinguishing Blast-Related Mild Traumatic Brain Injury from Posttraumatic Stress Disorder

One of the striking features of the MTBI cases being seen in the current OIF/OEF veterans is the high prevalence of posttraumatic stress disorder (PTSD). Indeed, PTSD or depression is probably present in most OIF/OEF veterans with suspected postconcussive syndromes secondary to MTBI.⁴ This coincidence could reflect dual exposure to both blasts and stressors that can independently cause PTSD. However, the clinical distinction between postconcussive syndrome and PTSD is often difficult, with the 2 disorders having many overlapping symptoms. For example, in both disorders, complaints of fatigue, irritability, poor

Table 1. *Important Components of the History in a Blast Injury.*

How far was the victim from the blast?
Did the victim experience a blast wave, and if so, was he thrown and did he strike another object as a result of this wave?
Was there a loss of consciousness?
Was there an altered level of consciousness (eg, seeing stars or feeling dazed)?
Was there amnesia surrounding the event, and if so, how long did it last?
Was there any additional trauma above the shoulders (eg, a blow to the head inside a vehicle exposed to the blast, lacerations, or loss of hearing)?
Was the veteran evacuated from the scene for medical treatment?

sleep, and impaired concentration, attention, and memory can be common. In practice, when a documented episode of TBI is present, the clinical distinction between the 2 disorders is typically based on the predominant symptoms: a postconcussive syndrome is suspected when patients have more organic symptoms such as headache, dizziness, visual complaints, hearing loss, balance problems, and cognitive disturbance, whereas PTSD is suspected when the predominant features are symptoms such as nightmares, hyperarousal, avoidance, and re-experiencing phenomena. The more severe cognitive features and associated neurological deficits make cases of moderate-to-severe TBI easy to recognize. What is more difficult is separating the postconcussive syndrome of MTBI from PTSD. Consequently, a detailed history is essential in making the diagnosis. However, sometimes even deciding whether a TBI has occurred may be difficult in cases of MTBI, in which distinguishing transient neurological dysfunction from a psychologically based reaction to stress is not always easy.

To date, the largest study addressing these issues is that of Hoge *et al.*,⁴ who surveyed over 2700 US Army infantry soldiers from 2 brigades 3 to 4 months after they returned from a 1-year deployment in Iraq. Soldiers responded to a questionnaire that elicited information regarding the occurrence of a TBI or other injury during deployment as well as their current general health status and the presence of symptoms suggestive of a postconcussion syndrome, PTSD, or depression. The definition of TBI used was an incident associated with LOC or altered mental status (ie, being dazed or confused, seeing stars, or not remembering the incident). The predominant TBI exposure was blast, and in the sample, 5% reported a TBI with LOC that typically lasted a few seconds to 3 minutes. Ten percent of the soldiers reported a TBI without LOC, and this meant a total of 15% who reported a TBI; among the 384 TBI cases, all but 4 were classified as mild. In comparison to soldiers who had suffered other types of physical injuries that did not involve TBI, soldiers with MTBI, particularly those who had experienced LOC, reported poorer general health, more missed workdays, and a higher number of somatic complaints. Complaints such as headache and memory and concentration problems were also common, suggesting that a persistent postconcussive syndrome was common in those who had suffered a MTBI. Of those reporting TBI with LOC, 44% met the criteria for PTSD, whereas PTSD was present in 27% of those reporting altered mental status without LOC. In addition, major depression was present in 23% and 8%, respectively. This high coincidence of PTSD and depression led the authors

to perform a covariate analysis for the 2 disorders, and interestingly, after adjustment for the coexistence of PTSD and depression, an MTBI history was no longer significantly associated with adverse physical health outcomes or symptoms, except for headache.

Given the lack of definitive biological markers for distinguishing TBI and PTSD, it is hard to know the exact meaning of these observations. One interpretation of these findings is that much of what is presently being called MTBI is really PTSD. In fact, similar suggestions have been made concerning the postconcussion syndrome following closed head injuries in civilian cases, with 1 recent prospective study reporting that postconcussive symptoms are as common following trauma without TBI as with MTBI.²²

The relationship between MTBI and PTSD is interesting because the 2 disorders can in one sense be considered different ends of a spectrum, with TBI being the classic example of an organic brain disease and PTSD being a psychologically based reaction to a stressor that was not associated with physical injury. Indeed, it has been suggested that the posttraumatic amnesia associated with TBI may in fact protect against PTSD, the argument being that if an individual cannot remember the incident, how can he have been psychologically traumatized by it, and 1 prospective study did find that PTSD rates were significantly higher in subjects who remembered the TBI incident versus those with no memory of the event.²³ However, because service personnel in a war zone inevitably have exposure to PTSD stressors independent of TBI events, the coexistence of the 2 disorders is easy to imagine.

The distinction between the 2 disorders has more than academic significance as it affects both treatment strategies and patient education. The pathophysiology of PTSD is most commonly conceptualized as an abnormally sustained stress response. Treatment of PTSD is focused on normalization of these reactions through psychologically based therapies as well as pharmacologically based treatments, the latter often involving the use of selective serotonin reuptake inhibitors. By contrast, TBI treatments are based on an organic model which presumes that structural brain alterations have occurred and that recovery depends on neurological factors, with treatments focused on improving attention and concentration with agents such as psychostimulants or improving compensatory strategies through cognitive/behavioral therapies.

A second possibility regarding the PTSD/TBI interface is that the 2 conditions are not coincidental but instead TBI may predispose individuals to developing PTSD. It is generally accepted that

individuals react to stressors differently and that maladaptive cognitive strategies such as prolonged rumination over an event or the belief that one will never recover is a predictor of subsequent PTSD.²⁴ Because even MTBI can impair cognitive reserve, it may compromise the ability to engage in positive cognitive strategies needed to manage the aftermath of a psychological trauma. Indeed, in an article accompanying that of Hoge *et al.*⁴ in the *New England Journal of Medicine*, Bryant²⁵ speculates that TBI might do this by damaging brain structures that are thought to be critical in PTSD. For example, TBI often damages the frontal lobes, and patients with PTSD have diminished medial prefrontal activation in reaction to fear.²⁶ Some biological models of PTSD postulate that a basic mechanism underlying PTSD is an exaggerated response of the amygdala during conditioned fear responses, and studies in humans and animals have suggested that modulation of these reactions involves inhibition of the amygdala by the medial prefrontal cortex.²⁶ Damage to the prefrontal cortex by TBI could therefore predispose individuals to abnormally sustained responses to psychological stressors in addition to impairing the cognitive reserve needed to deal with them afterward.

One last possibility is that we are witnessing a new and distinct cognitive and emotional dysregulation syndrome that combines features of both MTBI and PTSD. Perhaps MTBI predisposes to PTSD, and this in turn makes it difficult for the cognitively impaired brain to process PTSD symptoms.

TRAUMATIC BRAIN INJURY AND POSTTRAUMATIC STRESS DISORDER VIEWED FROM THE HISTORICAL PERSPECTIVE OF SHELL SHOCK

MTBI has been labeled the signature injury of the current wars in Iraq and Afghanistan. As discussed, a clear distinction between MTBI and PTSD is difficult; this problem is in fact not new and has roots dating back to the historical entity known as shell shock.²⁷ During World War I, while British troops were engaged in static trench warfare, which was characteristic of fighting on the frontlines during World War I, they were exposed to a variety of blasts at close range, including artillery barrages and mortar attacks. In this era before steel helmets were introduced, symptoms developed that were reminiscent of both the postconcussion syndrome and what would presently be called PTSD. A variety of labels for this entity were used, including the term *commotio cerebri*, but the most enduring label was

shell shock.²⁷ Shell shock became so common during World War I that 10% of British battle casualties carried the diagnosis, accounting for one-seventh of all discharges from the British Army and one-third of cases if war wounds were excluded. The pension claims associated with shell shock became enormously expensive for the British government. A vigorous debate took place concerning the nature of shell shock and whether it represented a physical injury or was the result of psychic trauma. The debate ended without any clear resolution but with probably most siding with a psychological explanation. As World War II began, hoping to avoid another epidemic, the British government went so far as to ban the use of the term *shell shock*. Despite this, soldiers continued to be exposed to blasts and to present with a similar range of symptoms. The controversy regarding physical causes versus psychological causes continued again without clear resolution. After the Vietnam era, PTSD became a defined entity, and there was less interest in blast-related TBI until the present time.

Clearly, elements of the shell shock controversy remain as a fundamental concern in military and disaster medicine. The distinction between TBI and PTSD has implications for treatment, patient education, and prognosis. The diagnosis of TBI may be viewed as carrying the implication of permanent brain injury, and labels can unintentionally affect prognosis. For example, 1 recent study of 73 patients admitted to an emergency room with MTBI found that negative beliefs regarding the likelihood of recovery were associated with poorer outcomes 3 months later.²⁸ By contrast, although the diagnosis of PTSD may be regarded as more benign in one sense, it still carries a stigma in the minds of some. This may be particularly true in the culture of the military, in which service personnel may be more comfortable and receptive to treatment if they are diagnosed as suffering from a TBI, a medical diagnosis, rather than PTSD, a psychological disorder. Interestingly, parallels can again be seen in the story of shell shock. Early in World War I, when shell shock was conceptualized as a form of war wound, the diagnosis was well accepted. Only later when it was suggested to be a form of neurosis or malingering did the diagnosis become controversial.²⁷

CONCLUSION

Key concepts concerning blast-related TBI are summarized in Table 2. A 2008 Rand Corporation study estimated that 20% of the first 1.6 million service personnel who served in OIF/OEF operations suffered a

Table 2. Key Concepts in Blast-Related Traumatic Brain Injury.

Blast injury is the most common cause of traumatic brain injury in the current military operations in Iraq and Afghanistan.
Blast-related injury occurs through the effects of the blast overpressure wave itself as well as the effects of objects being propelled by or individuals being thrown by the blast wind.
Studies in experimental animals have shown that a primary blast wave can induce brain injury.
It is unclear the degree to which primary blast waves damage the brain in humans, particularly in the setting of mild traumatic brain injury.
There is a high association of mild traumatic brain injury with posttraumatic stress disorder in returning Operation Iraqi Freedom/Operation Enduring Freedom veterans, and separating the 2 disorders clinically is often difficult.

TBI.¹ Of these, approximately 60% have never been assessed by a healthcare provider specifically for TBI. The exact costs for treating the identified cases is unknown. However, the Rand study estimated that through 2007, the cost of deployment-related TBI was between \$554 million and \$854 million. Final costs are unknown given the uncertainty over the length of the conflict as well as uncertainty over the natural history of blast-related MTBI. Estimates of the economic cost of PTSD run even higher. If blast-related TBI contributes to the development of mental health disorders, then the economic costs of TBI become enormous. In addition, nothing is known about the long-term consequences of blast-related TBI. All these issues have implications for the VA healthcare system. These uncertainties also have practical implications for military readiness, including issues such as identifying MTBI in the field, determining when to remove a soldier from the battle zone, and assessing the risk of redeployment once a TBI has been sustained. Resolution of these issues is likely to come about only through research into the basis of blast-related injury and its relationship to both cognitive and mental health.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Tanielian T, Jaycox LH. *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery*. Santa Monica, CA: Rand; 2008.

2. O'Hanlon ME, Campbell JH. Iraq index: Tracking variables of reconstruction and security in post-Saddam Iraq. Available at: <http://www.brookings.edu/saban/~media/files/centers/saban/iraq%20index/index20081120.pdf>. Accessed January 2009.

3. Warden DL, Ryan L, Helmick K, et al. War neurotrauma: the defense and veterans brain injury center (DVBIC) experience at the Walter Reed Army Medical Center. *J Neurotrauma* 2005; 22: 1178.

4. Hoge CW, McGurk D, Thomas JL, et al. Mild traumatic brain injury in U.S. soldiers returning from Iraq. *N Engl J Med* 2008; 358: 453–463.

5. Alvarez L. War veterans concussions are often overlooked. *New York Times* August 26, 2008; sect. A1.

6. Defense and Veterans Brain Injury Center. Available at: <http://www.dvbic.org>. Accessed January 2009.

7. VA Polytrauma System of Care Home. Available at: <http://www.polytrauma.va.gov>. Accessed January 2009.

8. Congressionally Directed Medical Research Programs. Available at: <http://cdmrp.army.mil>. Accessed January 2009.

9. DePalma RG, Burris DG, Champion HR, Hodgson MJ. Blast injuries. *N Engl J Med* 2005; 352: 1335–1342.

10. Taber KH, Warden DL, Hurley RA. Blast-related traumatic brain injury: what is known? *J Neuropsychiatry Clin Neurosci* 2006; 18: 141–145.

11. Chavko M, Koller WA, Prusaczyk WK, McCarron RM. Measurement of blast wave by a miniature fiber optic pressure transducer in the rat brain. *J Neurosci Methods* 2007; 159: 277–281.

12. Axelsson H, Hjelmqvist H, Medin A, et al. Physiological changes in pigs exposed to a blast wave from a detonating high-explosive charge. *Mil Med* 2000; 165: 119–126.

13. Chavko M, Prusaczyk WK, McCarron RM. Lung injury and recovery after exposure to blast overpressure. *J Trauma* 2006; 61: 933–942.

14. Cernak I, Savic J, Malicevic Z, et al. Leukotrienes in the pathogenesis of pulmonary blast injury. *J Trauma* 1996; 40(suppl): S148–S151.

15. Cernak I, Wang Z, Jiang J, et al. Cognitive deficits following blast injury-induced neurotrauma: possible involvement of nitric oxide. *Brain Inj* 2001; 15: 593–612.

16. Cernak I, Wang Z, Jiang J, et al. Ultrastructural and functional characteristics of blast injury-induced neurotrauma. *J Trauma* 2001; 50: 695–706.

17. Saljo A, Bao F, Haglid KG, Hansson HA. Blast exposure causes redistribution of phosphorylated neurofilament subunits in neurons of the adult rat brain. *J Neurotrauma* 2000; 17: 719–726.

18. Mochhala SM, Md S, Lu J, et al. Neuroprotective role of aminoguanidine in behavioral changes after blast injury. *J Trauma* 2004; 56: 393–403.

19. Gennarelli TA, Graham DI. Neuropathology. In: Silver JM, McAllister TW, Yudofsky SC, eds. *Textbook of Traumatic Brain Injury*. Arlington VA: American Psychiatric Publishing; 2005; 27–50.

20. Bhattacharjee Y. Neuroscience. Shell shock revisited: solving the puzzle of blast trauma. *Science* 2008; 319: 406–408.

21. Xydakis MS, Bebartha VS, Harrison CD, et al. Tympanic-membrane perforation as a marker of concussive brain injury in Iraq. *N Engl J Med* 2007; 357: 830–831.

22. Meares S, Shores FA, Taylor AJ, et al. Mild traumatic brain injury does not predict acute postconcussion syndrome. *J Neurol Neurosurg Psychiatry* 2008; 79: 300–306.
23. Gil S, Caspi Y, Ben-Ari IZ, et al. Does memory of a traumatic event increase the risk for posttraumatic stress disorder in patients with traumatic brain injury? A prospective study. *Am J Psychiatry* 2005; 162: 963–969.
24. Ehlers A, Clark DM. A cognitive model of posttraumatic stress disorder. *Behav Res Ther* 2000; 38: 319–345.
25. Bryant RA. Disentangling mild traumatic brain injury and stress reactions. *N Engl J Med* 2008; 358: 525–527.
26. Rauch SL, Shin LM, Phelps EA. Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research—past, present, and future. *Biol Psychiatry* 2006; 60: 376–382.
27. Jones E, Fear NT, Wessely S. Shell shock and mild traumatic brain injury: a historical review. *Am J Psychiatry* 2007; 164: 1641–1645.
28. Whittaker R, Kemp S, House A. Illness perceptions and outcome in mild head injury: a longitudinal study. *J Neurol Neurosurg Psychiatry* 2007; 78: 644–646.

Severe Traumatic Brain Injury: Maximizing Outcomes

Mary E. Tang, MD,¹ and Darlene A. Lobel, MD²

¹Department of Emergency Medicine, University of Florida, Jacksonville, FL

²Department of Neurosurgery, University of Florida, Jacksonville, FL

ABSTRACT

Severe traumatic brain injury is one of the leading causes of death and disability in the United States. The initial management of traumatic brain injury involves early resuscitation, computed tomography scanning, and surgical evacuation of mass lesions, when indicated. Recent research suggests that the prevention and treatment of secondary brain injury decrease mortality and improve outcomes. Specifically, treatment should address not only cerebral protection but also prevention of injury to other organ systems. To achieve the best outcomes, attention must be focused on optimizing blood pressure and brain tissue oxygenation, maintaining adequate cerebral perfusion pressures, and preventing seizures. In addition, maximizing good outcomes depends on proactively addressing the risk of common sequelae of brain injury, including infection, deep venous thrombosis, and inadequate nutrition. Guidelines developed for the management of severe traumatic brain injury have dramatically improved functional neurological outcomes. *Mt Sinai J Med* 76:119–128, 2009. © 2009 Mount Sinai School of Medicine

Key Words: brain injury, brain mass lesions, brain tissue oxygenation, cerebral perfusion pressures, computed tomography, injury sequelae, secondary injury, seizures, severe injury, traumatic injury.

Severe traumatic brain injury (TBI) is typically defined by a Glasgow Coma Scale (GCS) score of 3 to 8. Research has shown that not all injury occurs at the time of impact, with some injury occurring

over the hours and days after impact.¹ As a result, therapies aimed at preventing or managing this delayed injury can dramatically improve outcomes. Mortality studies have shown that some patients who initially present with a normal or relatively normal neurological examination will subsequently die. Thus, the importance of using serial examinations in evaluating patients with TBI is a cornerstone to management.¹

Ideal management of severe TBI patients begins with transport to a trauma center with neurosurgical capabilities and resuscitation based on meticulous airway and blood pressure management. Early intubation is recommended for severe TBI patients demonstrating poor airway protection or oxygenation. In urban settings, patients who do not require immediate intubation should receive supplemental oxygen and rapid transport. Furthermore, head computed tomography (CT), immediate surgical evacuation of intracranial mass lesions if indicated, and careful intensive care unit monitoring compose the mainstays of management. Brain swelling with associated increased intracranial pressure (ICP) develops in 10% to 15% of severe TBI patients who have initial normal head CT and in 53% to 63% of patients with acute traumatic abnormalities on admission CT.¹ Elevated ICP, along with the resulting decreased cerebral perfusion pressures (CPPs), is one of the most common causes of death and neurological disability.² Maximizing outcomes in severe brain injury requires attention to maintaining normal ICP, optimizing CPP, improving brain tissue oxygenation, and preventing other organ system injuries. The frequency of CT scans after head trauma is still a matter of debate. There is some evidence that shows repeating a CT scan at a standard time of 24 or 48 hours after TBI without a concomitant change in the patient's neurological status does not affect management and may worsen outcome by subjecting the patient to an unnecessary transfer to and from the CT scanner.³

The prevalence of TBI and its impact on society are staggering. There are 1.5 million reported cases

Address Correspondence to:

Darlene A. Lobel

Department of Neurosurgery

University of Florida

Jacksonville, FL

Email: darlene.lobel@jax.ufl.edu

of TBI annually in the United States; of these, 50,000 result in death.⁴ In cases of severe TBI, mortality estimates range from 30% to 50%. Furthermore, it is estimated that approximately 2% of the US population requires supportive care and assistance with activities of daily living as a result of TBI: this annual cost in the United States alone is \$100 billion.⁴ The incidence of mortality has decreased over the last 30 years because of changes in management based on data acquired from well-designed studies and from the application of evidence-based practice guidelines.¹ This article reviews the research and guidelines that have resulted in these improved outcomes.

METHODOLOGY

The majority of information provided in this article is based on the third edition of "Guidelines for the Management of Severe Traumatic Brain Injury," which was published by the Brain Trauma Foundation in 2007.¹ This is a highly regarded evidence-based practice guideline and is a definitive work on TBI. Inclusion criteria for studies reviewed in these guidelines were restricted to human studies published in the English language from 1966 to April 2006. Other inclusion criteria required adults at least 18 years old, a study population of at least 25, and in-hospital studies. Studies were restricted to randomized controlled trials, cohort and case-control studies, case series, databases, and registries. Studies in which more than 15% of the patients were pediatric or more than 15% of the patients had pathological processes other than TBI were excluded. Case reports, editorials, comments, and letters were also excluded. Studies that met inclusion criteria were independently reviewed by 2 members who assigned evidence classification. A larger board of reviewers then reviewed the classification of evidence for each study and determined the level of recommendation for each subject area. The final manuscript was then peer-reviewed.

According to "Guidelines for the Management of Severe Traumatic Brain Injury," evidence is classified as class I if it is supported by good-quality randomized controlled trials, class II if it is supported by cohort or case-controlled studies, and class III if it is supported by case series (Table 1). Likewise, levels of recommendation vary from level I, for which there is strong evidence for effectiveness, to level III, for which clinical certainty has not been established (Table 1).

In addition to the Brain Trauma Foundation recommendations, a PubMed literature search was completed of the literature published from April 2006

Table 1. Evidence Classification and Recommendation Levels According to the Brain Trauma Foundation Guidelines.

Classification of Evidence	
I	Good-quality randomized controlled trial
II	Moderate-quality randomized controlled trial, good-quality cohort study, or good-quality case-control study
III	Poor-quality randomized controlled trial, moderate/poor-quality cohort study, moderate/poor-quality case-control study, case series, database, or registry
Levels of Recommendation	
I	Strongest evidence for effectiveness
II	Moderate degree of clinical certainty
III	Clinical certainty not established

through November 2008; this included the search terms *traumatic brain injury*, *traumatic brain injury* and *steroids*, *traumatic brain injury* and *seizures*, *traumatic brain injury* and *critical care*, *traumatic brain injury* and *hypothermia*, *management of elevated intracranial pressure*, and *traumatic brain injury* and *brain oxygenation*.

BLOOD PRESSURE, CEREBRAL PERFUSION PRESSURE, AND CEREBRAL OXYGENATION

Maintaining adequate cerebral perfusion and brain tissue oxygenation remains one of the most critical facets of protection of the injured brain (see Table 2). Systemic hypotension and hypoxemia may produce secondary brain injury.⁵ Prevention of hypoxemia and inadequate brain oxygenation begins with early intubation in the prehospital setting, if indicated, and continues in the intensive care unit with monitoring of brain tissue oxygenation with specialized intracranial probes such as fiber-optic monitors. Hypoxemia, defined as oxygen saturations less than 90% or a partial pressure of arterial oxygen less than 60 mm Hg, has been shown to significantly increase mortality, even when these levels are sustained for only 11 to 20 minutes.⁶ Jones *et al.*⁶ conducted a subgroup analysis of 71 patients with secondary insults from a total of 124 inpatient TBI patients and found hypoxemia to be an independent predictor of mortality ($P = 0.024$). Recent research has shown that maintenance of adequate brain tissue oxygenation (>15 – 20 mm Hg) is associated with improved outcomes from cerebral trauma.⁷

It is equally as critical to avoid hypotension, which is defined as systolic blood pressure less than

Table 2. Summary of Recommendations To Optimize Cerebral Protection in Traumatic Brain Injury.

Cerebral Protection Parameter	Recommendations
Oxygenation	SaO ₂ > 90%, PaO ₂ > 60 mm Hg (level III)
Blood pressure	SBP > 90 mm Hg (level II)
Cerebral perfusion pressure	(MAP – ICP) > 50 mm Hg (level III)

Abbreviations: ICP, intracranial pressure; MAP, mean arterial pressure; PaO₂, partial pressure of arterial oxygen; SaO₂, saturation of arterial oxygen; SBP, systolic blood pressure.

90 mm Hg. Manley *et al.*⁸ evaluated the effect of hypotension in a prospective cohort study of 107 patients with a GCS score of less than 13 and found that 2 episodes of hypotension, even for a few minutes, increased the relative risk (RR) of mortality in TBI patients by 8 times. Furthermore, according to the Traumatic Coma Data Bank, a large prospectively collected data bank of over 1000 patients, a single episode of hypotension is among the 5 most significant predictors of outcome in severe TBI.^{9,10}

Likewise, CPP, measured as the difference between the ICP and mean arterial pressure (MAP; ie, CPP = MAP – ICP), plays an integral role in outcome from TBI. CPP should be maintained above 50 to 70 mm Hg H₂O to afford cerebral protection and decrease the risk of mortality.¹¹ As randomized controlled trials would, of course, be unethical in assessing the effects of hypotension and hypoxemia in TBI patients, the majority of this evidence falls into level II and level III recommendations, respectively.

Level II Recommendation

Blood pressure should be monitored in order to avoid systemic hypotension with systolic blood pressure less than 90 mm Hg.

Level III Recommendation

Oxygenation should be monitored in order to avoid hypoxia defined by a partial pressure of arterial oxygen less than 60 mm Hg or oxygen saturations less than 90%.

INTRACRANIAL PRESSURE

Sustained elevations in ICP significantly increase morbidity and mortality in TBI patients. Although normal

ICP is generally considered to be around 15 cm H₂O, the literature suggests that an ICP elevation greater than 20 cm H₂O for more than 5 minutes significantly affects outcome in TBI patients. The Brain Trauma Foundation reviewed 20 articles related to elevated ICP; of these, 10 were graded as level II, and it was concluded that patients whose ICP was maintained at 15 cm H₂O or lower had much better outcomes than those with ICPs of 20 to 25. It is therefore critical to monitor ICP in all salvageable patients with a GCS score of 3 to 8 and abnormal head CT (hematomas, contusions, cerebral edema, effacement of cisterns, or herniation). There is also some evidence to suggest that ICP monitoring should be performed in patients with severe TBI if 2 or more of these additional criteria are met: age greater than 40, systolic blood pressure less than 90 mm Hg, and posturing (Table 3).^{1,12}

ICP may be monitored with different devices, but commonly a fiber-optic monitor, placed in the parenchyma, or an external ventricular drain (EVD), placed in the frontal horn of the ventricle, is used to measure pressure.¹³ The benefit of the fiber-optic monitor is a continuous pressure measurement. An EVD provides the added benefit of treating elevated ICP by draining cerebral spinal fluid; however, the ICP level can be measured only when the drain is turned to the off position (ICP is measured through a fluid-coupled system). Another drawback of an EVD is the increased risk of hemorrhage associated with its placement due to the larger size of the catheter (5%) in comparison with the fiber-optic monitor (2%).¹⁴

Level II Recommendation

ICP monitoring should be performed in all patients with a possibility of survival with severe TBI and abnormal CT (swelling, hematoma, contusion, herniation, and compressed basal cisterns). Treatment should be initiated with ICP greater than 20 mm Hg,

Table 3. Management Strategies To Lower Elevated Intracranial Pressure Greater Than 20 cm H₂O.

Intervention	Mechanism
Head of bed greater than 30 degrees	Increase venous outflow
Mannitol and hypertonic normal saline	Diuretic (questionable in cerebral H ₂ O content)
Sedatives and analgesics	Cerebral metabolism/neuroprotective (questionable)
Barbiturates (only as a last resort)	Cerebral metabolism (questionable)
Surgery	Reduced mass effect
Steroids	No role, no benefit

but aggressive attempts to maintain CPP greater than 70 mm Hg with fluid and pressors should be avoided because of the risk of acute respiratory distress syndrome.

Level III Recommendation

ICP monitoring should be performed for severe TBI and normal CT if 2 or more of the following are present at admission: age greater than 40, unilateral/bilateral motor posturing, and systolic blood pressure less than 90. A combination of ICP values and clinical and CT findings is used to determine the need for treatment; CPP less than 50 mm Hg should be avoided.

CEREBRAL BLOOD FLOW AND HYPERVENTILATION

Management of elevated ICP is achieved through a variety of methods, both medical and surgical (see Table 4). One of the most rapidly responsive techniques is simply elevating the head of the bed by 30 degrees, which optimizes venous drainage from the brain. This effect is based on studies that assessed optimum cerebral blood flow and CPPs. It is important to remember that CPP can be compromised in this position by the lowering of MAP, and so it should be monitored carefully.¹⁵ Hyperventilation was a mainstay of ICP-lowering therapy for years. It reduces ICP by causing cerebral vasoconstriction and a subsequent reduction in cerebral blood flow. However, overly aggressive hyperventilation can lead to cerebral ischemia, and importantly, it must be noted that cerebral blood flow can be decreased in the hours following severe TBI independently of hyperventilation. As early as 1 day after severe TBI, cerebral blood flow diminishes to less than half that of a normal person. Thus, monitoring of the partial pressure of arterial carbon dioxide (PaCO₂) in these

TBI patients is integral to reducing the likelihood of cerebral ischemia from hypocapnia. Class II evidence shows that prophylactic hyperventilation to a PaCO₂ less than 25 mm Hg is harmful and thus not recommended.¹ As such, aggressive hyperventilation is recommended only as a temporizing measure for reduction of increased ICP and should be reversed as soon as there are no longer signs of brain herniation.

Level II Recommendation

Prophylactic hyperventilation to a PaCO₂ less than 25 is not recommended.

Level III Recommendation

Hyperventilation is recommended as a temporizing measure for reduction of increased ICP. This should be avoided, however, during the first 24 hours after injury when cerebral blood flow may be reduced or compromised.

DEHYDRATING AGENTS AND HYPEROSMOLAR THERAPY

As mentioned previously, drainage of cerebral spinal fluid via an EVD can aid in lowering ICP. Other management methods include use of diuretics, such as furosemide,¹⁶ to decrease cerebral blood volume, or mannitol, which exerts its effect via tissue dehydration while maintaining cerebral blood flow. Class II evidence suggests that mannitol at doses of 0.25 to 1 mg/kg is useful for treating elevated ICP as long as the systolic blood pressure is maintained at optimum levels (>90 mm Hg).

Hyperosmolar therapy, administered as 3% or 23% hypertonic normal saline (HTS), lowers ICP, most likely by the osmotic drive of water across the blood-brain barrier, which reduces the cerebral water content.^{1,17} HTS may also improve regional cerebral blood flow by a dehydration effect on swollen cerebrovascular endothelial cells.¹⁸ HTS should be cautiously administered in patients with preexisting hyponatremia because of the risk of developing central pontine myelinolysis.¹⁹ Central pontine myelinolysis is not seen in patients with normal sodium levels when HTS is given at the dose for ICP reduction.¹⁹ Furthermore, as HTS therapy causes an increase in intravascular volume, it can lead to increased cardiac output, so it should be used cautiously in patients with preexisting cardiopulmonary disease to avoid the development of pulmonary edema.²⁰

Table 4. Indications for Intracranial Pressure Monitoring According to Brain Trauma Foundation Guidelines.

Level II Recommendation	Level III Recommendation
GCS score of 3 to 8 and abnormal head CT	GCS score of 3 to 8 and 2 of the following: 1. Age >40 2. SBP <90 mm Hg 3. Posturing

Abbreviations: CT, computed tomography; GCS, Glasgow Coma Scale; SBP, systolic blood pressure.

Level II Recommendation

Mannitol is effective for the control of elevated ICP at doses of 0.25 to 1 g/kg. Hypotension with systolic blood pressure less than 90 mm Hg should be avoided.

Level III Recommendation

Mannitol use should be restricted prior to ICP monitoring to patients with signs of transtentorial herniation or progressive neurological deterioration not attributable to extracranial causes.

SEDATIVES

Sedatives and analgesics lower ICP by decreasing pain and agitation. However, it is important to note that high doses of these medications, particularly diprivan and propofol, can lead to complications, and so they should be used judiciously. A high-dose propofol infusion can lead to propofol infusion syndrome with possible clinical findings of hepatomegaly, hyperkalemia, metabolic acidosis, lipemia, myocardial dysfunction, renal failure, rhabdomyolysis, and death. Dosages greater than 5 mg/kg/hour in adults or administration for more than 48 hours should be used with caution.²¹

High-dose barbiturates have historically been used to lower ICP and provide cerebroprotective effects. However, there is no evidence that burst suppression via barbiturates should be administered prophylactically. This should be a last resort when all other medical and surgical management has failed. It is critical to maintain hemodynamic stability during barbiturate therapy.^{1,15} Finally, when a mass lesion, such as a hematoma, is present, surgery to evacuate the mass or to decompress a focal area of brain swelling is an important consideration in the management strategy.

Level II Recommendation

Prophylactic barbiturate usage to induce burst suppression on electroencephalogram is not recommended. High-dose barbiturates are recommended to control increased ICP refractory to maximum standard medical and surgical treatment.

STEROIDS

Steroids have no role in the treatment of TBI. The Brain Trauma Foundation reviewed 6 studies that

showed no clinical benefit from the administration of glucocorticoids in TBI treatment. One of these pivotal clinical trials, the Corticosteroid Randomisation After Significant Head Injury trial, was a randomized controlled trial that was actually halted early because of the significant increase in mortality in the group randomized to high-dose steroids versus the control group.²²

The theory behind steroid use is based on the inflammatory response in the injured brain, which leads to cerebral edema. Empirical evidence suggests that cerebral edema surrounding brain tumors responds to treatment with steroids. This evidence, in combination with animal studies showing that steroids lead to a decrease in vascular permeability, would suggest that steroids may provide a benefit in the treatment of edema in severe TBI. However, steroids have been shown definitively to be ineffective in decreasing elevated ICP, and the administration of methylprednisolone has been shown to actually increase mortality in TBI patients because the infection threshold is lowered.¹

Level I Recommendation

The use of steroids is not recommended.

HYPOTHERMIA

Only class III data exist on the topic of therapeutic hypothermia in the management of TBI.¹ Although prophylactic hypothermia is often initiated at admission and used to decrease cerebral oxygen demands, there is no conclusive scientific evidence that its usage decreases morbidity and mortality. In a meta-analysis conducted by the Brain Trauma Foundation, studies demonstrating class II evidence were examined to determine all-cause mortality and secondary outcomes such as favorable neurological status.¹ Specific aspects examined with respect to hypothermia included the target cooling temperature (32°C–33°C or >33°C), cooling duration (<48 hours, 48 hours, or >48 hours), and rate of rewarming (1°C/hour, 1°C/day, or slower). Primary data analysis from 6 studies that met inclusion criteria, which included 354 hypothermic patients versus 340 controls, showed that all-cause mortality was not significantly different in hypothermic patients versus controls (RR, 0.76; $P = 0.18$). Hypothermic patients did have a 46% increased chance of good outcome versus controls as measured by a Glasgow Outcome Scale (GOS) score of 4 or 5 (RR, 1.46; $P = 0.006$). Subgroup analyses showed decreased mortality when cooling was maintained for greater than 48 hours [RR, 0.51; 95%

confidence interval (CI), 0.34–0.78]. No effect on mortality was found that was related to the target cooling temperature or rate of rewarming. Improved outcomes with a GOS score of 4 or 5 were found in groups achieving target temperatures of 32°C to 33°C (RR, 1.67; 95% CI, 1.18–2.35) and 33°C to 35°C (RR, 1.75; 95% CI, 1.12–2.73). There was no clear improvement in outcomes with respect to cooling duration or rate of rewarming. These results suggest that although therapeutic hypothermia may improve outcomes in patients who survive and may reduce mortality caused by neurological injury, overall, the all-cause mortality rate is not affected by this therapy.

Level III Recommendation

In pooled data, prophylactic hypothermia as a treatment modality is not significantly associated with mortality in comparison with normothermic controls. A greater decrease in the mortality rate is achieved when target hypothermic temperatures are maintained for more than 48 hours. Prophylactic hypothermia is associated with significantly higher GOS scores in comparison with normothermic patients.

SEIZURE PROPHYLAXIS

Posttraumatic seizures are categorized as either early (within 7 days of injury) or late (greater than 7 days post-injury). Incidence of seizures after penetrating injuries is approximately 50% in patients followed for 15 years.²³ The incidence of early posttraumatic seizures in severe TBI ranges from 4% to 25%, whereas the incidence of late posttraumatic seizures is 9% to 42% in untreated patients.²⁴ Risk factors for early posttraumatic seizures include a GCS score less than 10, cortical contusion, depressed skull fracture, subdural hematoma, epidural hematoma, intracerebral hematoma, penetrating head injury, and seizure within 24 hours of injury. Although no direct outcome studies exist, it is generally agreed that it is desirable to avoid seizures in the early period to prevent further insults to the brain due to increased ICP, blood pressure and oxygen delivery changes, and excess neurotransmitter release. Recent research suggests that prevention of early posttraumatic seizures may also prevent the development of chronic epilepsy.^{1,24} Therefore, anti-epileptic drugs (AEDs) are recommended to aid in the prevention of early posttraumatic seizures within 7 days of injury when the aforementioned risk factors are met.

There is no evidence to suggest that continuing prophylactic administration of AEDs after the first

7 days of injury improves outcome or reduces long-term seizure risk.²⁵ The American Academy of Neurology (AAN) practice parameters on AED prophylaxis in severe TBI also support these statements.²⁶ The AAN's practice guidelines recommend that for adult patients with severe TBI, prophylactic treatment with phenytoin, starting with an intravenous loading dose, be administered as soon as possible post-injury to decrease the risk of early posttraumatic seizures (level A recommendation). With a lesser degree of strength of evidence, the AAN recommends that AEDs should not be routinely started after the first 7 days to decrease the risk of late posttraumatic seizures (level B). Furthermore, even patients who do have 1 seizure in the first week after injury are not candidates for long-term AED therapy. Those patients who experience 2 or more seizures in the early post-traumatic period should be assessed as new-onset seizure patients and considered for AED treatment.²⁷

Most studies do not support the use of anti-epileptics for late posttraumatic seizures. If late posttraumatic seizures occur, patients should be managed as new-onset seizure patients. Patients who do not develop seizures in the acute TBI phase have better outcomes than those patients who do have seizures.²⁴ Interestingly, the poorest outcomes are seen in patients who undergo surgery and develop late-onset seizures during the acute phase of TBI.¹ Historically, the AED of choice for the treatment of early posttraumatic seizures has been phenytoin, which has documented clinical efficacy in reducing the incidence of early posttraumatic seizures.¹ Recently, levetiracetam has demonstrated potential benefit.²⁸

Level II Recommendation

Prophylactic use of phenytoin or valproate is not recommended for prevention of late posttraumatic seizures (after 7 days). Anti-epileptics are recommended to aid in the prevention of early posttraumatic seizures within 7 days of injury.

PROPHYLAXIS

Nutrition

Only a few studies exist that have evaluated the impact of nutrition on recovery from TBI.^{29–31} Recent literature has clearly disproved the popular belief that a patient in a coma has a lower metabolic rate and therefore a lower caloric intake requirement. On the contrary, what is actually seen is an increase of metabolism by, on average, 140% in these critically

ill patients.³⁰ Paralyzed patients, interestingly, have comparatively decreased metabolic requirements (even though these requirements are still elevated above those of normal patients), and this suggests that the hypermetabolism and nitrogen wasting seen may be due to increased muscle tone.³² Outcome studies have shown that poor nutrition over a 2-week period increases the mortality rate in comparison with those patients who are maintained on their full caloric levels by 7 days after injury.³³ Therefore, level II recommendations state that by postinjury day 7, patients should be nourished to maintain full caloric intake.¹

Hyperglycemia (capillary blood glucose >110 mg/dL) has been shown to exacerbate hypoxic-ischemic injury, and sustained hyperglycemic levels are associated with poorer outcomes. In a retrospective review of 169 patients with moderate or severe TBI, Lam *et al.*³⁴ found that severe TBI patients with serum glucose greater than 200 mg/dL had significantly poorer outcomes as measured by the GOS ($P < 0.01$). Insulin drips may be used to maintain normal glucose levels in head-injured patients. However, it is critical to maintain tight control of the glucose levels (80–110 mg/dL) as this is the brain's primary source of energy and is critical to the healing process of the injured brain.

Infection Prevention

In patients with TBI, a higher incidence of infection with mechanical ventilation and invasive monitoring techniques leads to increased length of stay and higher mortality. In a randomized controlled trial of 100 mechanically ventilated intensive care unit patients, 86% with severe TBI, Sirvent *et al.*³⁵ administered 2 doses of cefuroxime (1.5 g) within 6 hours after intubation in a treatment group versus no postintubation antibiotics for the control group. The resultant incidence of pneumonia was 23% versus 64% in the treatment and control groups, respectively ($P = 0.016$). Early tracheostomy is also recommended as it decreases the days of mechanical ventilation.¹ Interestingly, however, early tracheostomy does not decrease mortality or the rate of nosocomial pneumonia.^{36,37} Sugerman *et al.*³⁷ conducted a multicenter randomized controlled trial with 127 patients, 67 of whom had severe TBI. Thirty-five patients underwent early tracheostomy on days 3 to 5, whereas 32 patients had late tracheostomy performed on days 10 to 14; there was no significant difference in the incidence of pneumonia or death.

Systemic nosocomial infection rates also increase with TBI severity and coexisting chest trauma.³⁸ Prolonged courses of prophylactic antibiotics do not

show a reduction in nosocomial infections. In a study by Goodpasture *et al.*³⁹ although prophylactic antibiotics reduced the incidence of postintubation pneumonia, the patients treated with antibiotics were noted to have earlier appearance of severe infections with gram-negative organisms. Patients who were not treated with antibiotics developed less severe infections with gram-positive organisms.

The incidence of infection with ICP monitor usage ranges from less than 1% up to 27%.⁴⁰ Infection depends on several factors, including duration of monitoring, presence of systemic infection, use of prophylactic intravenous antibiotics, presence of open skull fracture, intraventricular hemorrhage or subarachnoid hemorrhage, leakage around the ventriculostomy catheter, and flushing of ventriculostomy tubing. Routine catheter exchange to prevent cerebrospinal fluid infection is not supported by level III evidence.^{1,40}

Level II Recommendation

Periprocedural antibiotics for intubation should be given to decrease the incidence of pneumonia. This does not change the mortality or length of stay. Early tracheostomy is beneficial to decrease days of mechanical ventilation but does not decrease mortality or the rate of nosocomial pneumonia.

Level III Recommendation

Routine ventricular catheter exchange or prophylactic antibiotics for ventricular catheter placement are not recommended. Early extubation in patients with adequate gag and cough reflex does not increase the risk of pneumonia.

Deep Venous Thrombosis Prevention

According to data from the National Trauma Databank, severe TBI has been found to be a high risk factor for the development of venous thromboembolism (VTE), which includes deep venous thrombosis (DVT) and pulmonary embolism. Knudson *et al.*⁴¹ examined 1602 patients with VTE and found the total incidence to be 0.36% after trauma. Severe TBI was further found to be 1 of 6 independent risk factors for the development of VTE (odds ratio, 2.59; 95% CI, 2.31–2.90). In addition to the immobility associated with TBI, the ensuing hypercoagulable state that occurs after brain injury contributes to the high incidence of DVT in this patient population.⁴² DVT incidence without prophylaxis after severe TBI is approximately 20%.⁹ Page *et al.*⁴³ examined data from the Pennsylvania

Trauma Outcomes Study and found the incidence of pulmonary embolism in TBI patients to be 0.38%. There was no significant difference in pulmonary embolism incidence in TBI patients compared to those without head injury. Therefore, graduated compression stockings or intermittent pneumatic compression devices to reduce the incidence of DVT are recommended until patients are ambulatory (level III recommendation).¹

Pharmacological DVT prophylaxis holds more controversy. Agnelli *et al.*⁴⁴ found that in patients undergoing elective neurosurgical procedures, there was no significant increase in the rate of bleeding for patients given enoxaparin versus mechanical DVT prophylaxis.⁴⁴ Lower rates of DVT were found in the enoxaparin group compared to the mechanical prophylaxis group: 17% versus 32% ($P = 0.004$). However, some studies show that low-molecular-weight heparin or unfractionated heparin used with mechanical prophylaxis increases the risk of expansion of intracranial hemorrhage in the first 24 hours after initial injury.⁴⁵ There is insufficient evidence to recommend the preferred agent, dosage, timing, or specific pharmacological prophylaxis for DVT. Furthermore, data evaluating the risk versus benefit of pharmacological DVT prophylaxis in the setting of severe TBI are limited. Therefore, the decision regarding if and when to start anticoagulation after TBI is often made on a case-by-case basis.

Level III Recommendation

Graduated compression stockings or intermittent pneumatic compression stockings are recommended.

IMPACT OF GUIDELINES

The impact of the Brain Trauma Foundation's "Guidelines for the Management of Severe Traumatic Brain Injury" has been significant, with improved outcomes and potential for massive savings of healthcare costs. Fakhry *et al.*⁴⁰ examined data collected from a level I trauma center and found that using these guidelines decreased intensive care unit and total hospital length of stay by 1.8 and 5.4 days, respectively, resulting in a savings of \$4.7 million in hospital charges. Furthermore, mortality decreased by 4.0%, although this was not statistically significant, and improved outcomes as measured by the GOS score were found. During times of high compliance with the guidelines, 61.5% of severe TBI patients had a GOS score of 4 or 5 versus only 43.3% before guideline implementation ($P < 0.001$).

Table 5. *Management of Severe Traumatic Brain Injury: Key Concepts.*

Maintain SBP >90 mm Hg and PaO ₂ >60 mm Hg.
Treat increased ICP >20 mm Hg and maintain CPP at 50 to 70 mm Hg.
Avoid hyperventilation.
Steroids are not indicated in the treatment of traumatic brain injury.
Administer antiseizure prophylaxis to prevent early posttraumatic seizures.
Attention to prevention of infection, deep venous thrombosis, and poor nutrition is critical to traumatic brain injury care.

Abbreviations: CPP, cerebral perfusion pressure; ICP, intracranial pressure; PaO₂, partial pressure of arterial oxygen; SBP, systolic blood pressure.

In a web-based survey, Hesdorffer and Ghajar⁴⁷ determined that adherence to the Brain Trauma Foundation guidelines was higher in level I trauma centers and in facilities that responded by having treatment protocols in place. In addition, routine ICP monitoring and avoidance of steroids increased significantly from the time of the development of the Brain Trauma Foundation guidelines. Lastly, Faul *et al.*⁴⁸ determined that widespread implementation of the Brain Trauma Foundation guidelines would cost \$61 million, but the savings would far outweigh the cost. It was estimated that savings would amount to \$262 million in annual medical costs, \$43 million in annual rehabilitation costs, and \$3.84 billion in lifetime societal costs.

CONCLUSIONS

Severe TBI and the impact of secondary injury cause significant morbidity and mortality in the United States. Since the development of the 1st edition of the Brain Trauma Foundation's "Guidelines for the Management of Severe Traumatic Brain Injury" in 1995, outcomes after TBI have improved significantly. Critical to the prevention of secondary injury is the maintenance of adequate cerebral perfusion, brain tissue oxygenation, and systemic blood pressure. Monitoring of ICP should be performed in select patients, and ICP levels greater than 20 cm H₂O should be treated with furosemide, mannitol, or hypertonic saline and possibly EVD placement, if indicated. Sedatives and analgesics may also aid in lowering ICP. Steroids continue to be contraindicated in severe TBI.

Hypothermia is recommended to increase chances of improved outcome in severe TBI. Prevention of early posttraumatic seizures should involve

the administration of phenytoin or levetiracetam as soon as possible. AED prophylaxis for late post-traumatic seizures is not supported. With respect to infection prophylaxis, although a short course of antibiotics is recommended for the postintubation period, prolonged courses are not recommended. Furthermore, full nutrition should be achieved by postinjury day 7, and mechanical DVT prophylaxis should be administered in all patients without contraindications (Table 5). Acceptance of and adherence to the guidelines by level I trauma centers and other receiving facilities have dramatically improved outcomes for severely brain-injured patients.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Brain Trauma Foundation, American Association of Neurological Surgeons, Joint Section of Neurotrauma and Critical Care. Guidelines for the management of severe traumatic brain injury, 3rd ed. *J Neurotrauma* 2007; 24: S1–S106.
- Piek J, Chesnut RM, Marshall LF, et al. Extracranial complications of severe head injury. *J Neurosurg* 1992; 77: 901–907.
- Lee TT, Aldana PR, Kirton OC, et al. Follow-up computerized tomography (CT) scans in moderate and severe head injuries: correlation with Glasgow Coma Scores (GCS) and complication rate. *Acta Neurochir (Wien)* 1997; 139: 1042–1048.
- Rutland-Brown W, Langlois JA, Thomas KE, Xi YL. Incidence of traumatic brain injury in the United States, 2003. *J Head Trauma Rehabil* 2006; 21: 544–548.
- Stochetti N, Furlan A, Volta F. Hypoxemia and arterial hypotension at the accident scene in head injury. *J Trauma* 1996; 40: 764–767.
- Jones PA, Andrews PJD, Midgely S, et al. Measuring the burden of secondary insults in head injured patients during intensive care. *J Neurosurg Anesthesiol* 1994; 6: 4–14.
- Valadka AB, Robertson CS. Surgery of cerebral trauma and associated critical care. *Neurosurgery* 2007; 61: 203–221.
- Manley G, Knudson M, Morabito D, et al. Hypotension, hypoxia, and head injury: frequency, duration, and consequences. *Arch Surg* 2001; 136: 1118–1123.
- Chestnut RM, Marshall LF, Klauber MR, et al. The role of secondary brain injury in determining outcome from severe head injury. *J Trauma* 1993; 34: 216–222.
- Marmarou A, Anderson RL, Ward JD, et al. Impact of ICP instability and hypotension on outcome in patients with severe head trauma. *J Neurosurg* 1991; 75: 159–166.
- Rosner MJ, Rosner SD, Johnson AH. Cerebral perfusion pressure: management protocol and clinical results. *Neurosurgery* 1995; 83: 949–962.
- Narayan R, Kishore P, Becker D, et al. Intracranial pressure: to monitor or not to monitor? A review of our experience with head injury. *J Neurosurg* 1982; 56: 650–659.
- Zhong J, Dujovny M, Park HK, et al. Advances in ICP monitoring techniques. *Neurol Res* 2003; 25: 339–350.
- Kakarla UK, Kim LJ, Chang SW, et al. Safety and accuracy of bedside external ventricular drain placement. *Neurosurgery* 2008; 63(suppl 1): 162–166.
- Pickard JD, Czosnyka M. Management of raised intracranial pressure. *J Neurol Neurosurg Psychiatry* 1993; 56: 845–858.
- Woster PS, LeBlanc KL. Management of elevated intracranial pressure. *Clin Pharm* 1990; 9: 762–772.
- Qureshi AI, Suarez JJ, Castro A, et al. Use of hypertonic saline/acetate infusion in the treatment of cerebral edema in patients with head trauma: experience at a single center. *J Trauma* 1999; 47: 659–665.
- Qureshi AI, Suarez JJ. Use of hypertonic saline solutions in treatment of cerebral edema and intracranial hypertension. *Crit Care Med* 2000; 28: 3301–3313.
- Kleinschmidt-DeMasters BK, Norenberg MD. Rapid correction of hyponatremia causes demyelination: relation to central pontine myelinolysis. *Science* 1981; 211: 1068–1070.
- Qureshi AI, Suarez JJ, Bhardwaj A, et al. Use of hypertonic (3%) saline/acetate infusion in the treatment of cerebral edema: effect on intracranial pressure and lateral displacement of the brain. *Crit Care Med* 1998; 26: 440–446.
- Kang TF. Propofol infusion syndrome in critically ill patients. *Ann Pharmacother* 2002; 36: 1453–1456.
- Roberts I, Yates D, Sandercock P, et al. Effect of intravenous corticosteroids on death within 14 days in 10,008 adults with clinically significant head injury (MRC CRASH trial): randomized placebo-controlled trial. *Lancet* 2004; 364: 1321–1328.
- Temkin NR, Dikmen SS, Winn HR. Posttraumatic seizure. In: Eisenberg HM, Aldrich EF, eds. *Management of Head Injury*. Philadelphia, PA: WB Saunders; 1991; 425–435.
- Wang HC, Chang WN, Chang HW, et al. Factors predictive of outcome in posttraumatic seizures. *J Trauma Inj Infect Crit Care* 2008; 64: 883–888.
- Lossius MI, Hessen E, Mowinckel P, et al. To stop or not stop the AED? *Epilepsy Curr* 2008; 4: 90–91.
- Chang BS, Lowenstein DH. Practice parameter: antiepileptic drug prophylaxis in severe traumatic brain injury: report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2003; 60: 10–16.
- French JA, Kanner AM, Bautista J, et al. Efficacy and tolerability of the new antiepileptic drugs, I: Treatment of new-onset epilepsy: report of the TTA and QSS Subcommittees of the American Academy of Neurology and the American Epilepsy Society. *Epilepsia* 2004; 45: 401–409.
- Jones KE, Puccio AM, Harshman KJ, et al. Levetiracetam versus phenytoin for seizure prophylaxis in severe traumatic brain injury. *Neurosurg Focus* 2008; 25: E3.
- Borzotta AP, Pennings J, Papasadero B, et al. Enteral versus parenteral nutrition after severe closed head injury. *J Trauma* 1994; 37: 459–468.

30. Deutschman CS, Konstantinides FN, Raup S. Physiological and metabolic response to isolated closed-head injury. Part 1: Basal metabolic state: correlations of metabolic and physiological parameters with fasting and stressed controls. *J Neurosurg* 1986; 66: 89–98.
31. Young B, Ott L, Twyman D, et al. The effect of nutritional support on outcome from severe head injury. *J Neurosurg* 1987; 67: 668–676.
32. Clifton GL, Robertson CS, Choi SC. Assessment of nutritional requirements of head-injured patients. *J Neurosurg* 1986; 66: 895–890.
33. Rapp RP, Young B, Twyman D, et al. The favorable effect of early parental feeding on survival in head-injured patients. *J Neurosurg* 1983; 58: 906–912.
34. Lam AM, Winn HR, Cullen BF, et al. Hyperglycemia and neurological outcome in patients with head injury. *J Neurosurg* 1991; 75: 545–551.
35. Sirvent JM, Torres A, Mustafa E, et al. Protective effect of intravenously administered cefuroxime against nosocomial pneumonia in patients with structural coma. *Am J Respir Crit Care Med* 1997; 155: 1729–1734.
36. Ahmed N, Kuo YH. Early versus late tracheostomy in patients with severe traumatic head injury. *Surg Infect* 2007; 8: 343–347.
37. Sugerman HJ, Wolfe L, Pasquale MD, et al. Multicenter, randomized prospective trial of early tracheostomy. *J Trauma* 1997; 43: 741–747.
38. Helling TS, Evans LL, Fowler DL, et al. Infectious complications in patients with severe head injury. *J Trauma* 1988; 28: 1575–1577.
39. Goodpasture HC, Romig DA, Voth DW. A prospective study of tracheobronchial bacterial flora in acutely brain-injured patients with and without antibiotic prophylaxis. *J Neurosurg* 1977; 47: 228–235.
40. Dasic D, Hanna SJ, Bojanic S, et al. External ventricular drain infection: the effect of a strict protocol on infection rates and a review of the literature. *Br J Neurosurg* 2006; 20: 296–300.
41. Knudson MM, Ikosi DG, Khaw L, et al. Thromboembolism after trauma: an analysis of 1602 episodes from the American College of Surgeons National Trauma Data Bank. *Ann Surg* 2004; 240: 490–496.
42. Carlile MC, Yablon SA, Mysiwn WJ, et al. Deep venous thrombosis management following traumatic brain injury: a practice survey of the traumatic brain injury model systems. *J Head Trauma Rehabil* 2006; 21: 483–490.
43. Page RB, Spott MA, Krishnamurthy S, et al. Head injury and pulmonary embolism: a retrospective report based on the Pennsylvania Trauma Outcomes study. *Neurosurgery* 2004; 54: 143–148.
44. Agnelli G, Piovella F, Buoncristiani P, et al. Enoxaparin plus compression stocking compared with compression stocking alone in the prevention of venous thromboembolism after elective neurosurgery. *N Engl J Med* 1998; 339: 80–85.
45. Norwood SH, McAuley CE, Berne JD, et al. Prospective evaluation of enoxaparin prophylactics for venous thrombo-embolism in patients with intracranial hemorrhagic injuries. *Arch Surg* 2002; 137: 696–701.
46. Fakhry SM, Trask AL, Waller MA, et al. Management of brain-injured patients by an evidence-based medicine protocol improves outcomes and decreases hospital charges. *J Trauma* 2004; 56: 492–500.
47. Hesdorffer DC, Ghajar J. Marked improvement in adherence to traumatic brain injury guidelines in United States trauma centers. *J Trauma* 2007; 63: 841–848.
48. Faul M, Wald M, Rutland-Brown W, et al. Using a cost-benefit analysis to estimate outcomes of a clinical treatment guideline: testing the Brain Trauma Foundation Guidelines for the treatment of severe traumatic brain injury. *J Trauma* 2007; 63: 1271–1278.

Mild Traumatic Brain Injury

John J. Bruns Jr, MD and Andy S. Jagoda, MD

Department of Emergency Medicine, Mount Sinai School of Medicine, New York, NY

ABSTRACT

Mild traumatic brain injury accounts for 1% to 2% of emergency department visits in the United States. Up to 15% of these patients will have an acute intracranial lesion identified on head computed tomography; less than 1% of mild traumatic brain injuries will require neurosurgical intervention. Clinical research over the past decade has focused on identifying the subgroup of patients with mild traumatic brain injury with acute traumatic lesions on computed tomography and specifically those at risk for harboring a potentially catastrophic lesion. This research has been used to generate evidence-based guidelines to assist in clinical decision making. There is no evidence to support the use of plain film radiographs in the evaluation of patients with mild traumatic brain injury. The utility of brain-specific biomarkers is rapidly evolving, and a growing body of evidence supports their potential role in determining the need for neuroimaging. Clinical predictors for identifying patients with abnormal computed tomography have been established and, if used, may have a significant positive impact on traumatic brain injury–related morbidity and healthcare utilization in the United States. Patients with negative computed tomography are at almost no risk of deteriorating; however, they should be counseled regarding postconcussive symptoms and should be given appropriate written instructions and referrals at discharge. *Mt Sinai J Med* 76:129–137, 2009. © 2009 Mount Sinai School of Medicine

Key Words: brain-specific biomarkers, computed tomography, mild brain injury, postconcussive symptoms, traumatic brain injury.

Address Correspondence to:

John J. Bruns Jr, MD

Department of Emergency
Medicine

Mount Sinai School of Medicine
New York, NY

Email: john.bruns@mssm.edu

About 1.4 million incidents of traumatic brain injury (TBI) are reported in the United States each year,¹ of which 75% are typically classified as mild.² Most of these injuries result from falls and motor vehicle crashes. Falls are responsible for a significant proportion of mild traumatic brain injury (MTBI) at the extremes of age, and motor vehicle collisions are the predominant etiology in adolescent and young adult males.^{3,4} Up to 15% of patients with head trauma evaluated in the emergency department (ED) with a Glasgow Coma Scale (GCS) score of 15 will have acute traumatic pathology identified on head computed tomography (CT); however, less than 1% of these patients will require neurosurgical intervention.^{5–7} Depending on the criteria used to define disability, up to 15% of patients with MTBI will have compromised function 1 year after their injury.^{8,9} The primary objective in the initial evaluation of MTBI patients is to determine who has an acute traumatic intracranial injury and who can be safely discharged from the acute care setting.

DEFINITIONS

There is no consensus definition of MTBI. MTBI describes a state in which there is an alteration of minimal duration and severity or no change from the patient's baseline neurological or mental status after an injury. MTBI may not be clinically evident during the initial medical and neurological evaluation because of symptom resolution or the employment of insensitive assessments. Although the term *head injury* is often used interchangeably with *TBI*, it is an antiquated and an inappropriate synonym for *TBI*. *Head injury* is defined as clinically evident trauma above the clavicles, such as lacerations, ecchymosis, and forehead abrasions. *TBI* refers to perturbations of the brain's function itself and may occur in the absence of perceptible head injury. It is manifest as confusion, focal neurological deficits, an altered level of consciousness, and/or subtle abnormalities on neuropsychological examinations. *Concussion* is another term that is used interchangeably with *MTBI*,

Table 1. *Concussion Grading Scales and Return-To-Activity Recommendations.*

Concussion Severity	First Concussion	Second Concussion	Third Concussion
Grade 1 (mild) No LOC PTA < 30 minutes	May return to play if asymptomatic	May return in 2 weeks if asymptomatic at that time for 1 week	Should terminate season and may return next year if asymptomatic
Grade 2 (moderate) LOC < 5 minutes PTA > 30 minutes	May return after being asymptomatic for 1 week	Should wait at least 1 month, may return then if asymptomatic for 1 week, and should consider terminating season	Should terminate season and may return next year if asymptomatic
Grade 3 (severe) LOC > 5 minutes PTA > 24 hours	Should wait at least 1 month and may return then if asymptomatic for 1 week	Should terminate season and may return next year if asymptomatic	

Abbreviations: LOC, loss of consciousness; PTA, posttraumatic amnesia.

often in the context of sports. Concussion is often classified into 3 arbitrary grades (see Table 1¹⁰).

Grade I concussion has no associated loss of consciousness or amnesia, grade II has amnesia but no loss of consciousness, and grade III is associated with loss of consciousness.¹⁰ The system most often used for grading the severity of brain injury is the GCS.¹¹ An MTBI has historically been ascribed a score of 13 to 15. However, recent trends based on the higher association of positive CT, need for neurosurgical intervention, and long-term sequelae has resulted in an evolving consensus to exclude a GCS score of 13 from the mild category. The Centers for Disease Control and Prevention has developed a conceptual definition for MTBI,¹² that is, the occurrence of injury to the head resulting from blunt trauma or acceleration or deceleration forces with 1 or more of the following conditions attributable to the head injury during the surveillance time period:

- Any period of observed or self-reported transient confusion, disorientation, or impaired consciousness.
- Any period of observed or self-reported dysfunction of memory (amnesia) around the time of injury.
- Observed signs of other neurological or neuropsychological dysfunction.
- Any period of observed or self-reported loss of consciousness lasting 30 minutes or less.

The primary challenge to the ED physician is to detect the well appearing, neurologically normal MTBI patient with a potentially significant intracranial injury. A secondary challenge is to identify those patients at risk for having prolonged postconcussive symptoms and those at risk for the postconcussive syndrome (PCS) in order to anticipate and provide proper discharge instructions and referrals. The management of minor closed head injury in children

is presented in a clinical policy by the American Academy of Pediatrics and the American Academy of Family Physicians¹³; this article specifically concerns the management of MTBI in patients without multisystem injury, 16 years old or older, with a GCS score of 14 or 15 presenting within 24 hours of injury.

METHODOLOGY

The framework for this document is derived from the evidentiary tables created by the Center for Disease Control (CDC) and the American College of Emergency Physicians (ACEP) MTBI practice guideline task force.¹⁴ The evidentiary table was created after a review and critical analysis of MEDLINE and the Cochrane Database, which was searched for articles from January 2000 through 2007. Search result articles and their bibliographies were screened by panel members for relevance. All selected articles that were used were graded on the basis of applicability to the clinical questions. Articles received a final grade of class I, II, or III on the basis of a predetermined formula taking into account the design and quality of the articles. The strength of the recommendations regarding patient management was then determined according to the following criteria:

- Level A: Generally accepted principles for patient management that reflect a high degree of clinical certainty.
- Level B: Recommendations for patient management that reflect moderate clinical certainty.
- Level C: Other patient management strategies based on preliminary or inconclusive data or, in the absence of any published literature, panel consensus.

SKULL RADIOGRAPHY IN MILD TRAUMATIC BRAIN INJURY

Even though they are readily available and relatively inexpensive, numerous studies have shown that skull radiographs are neither sensitive nor specific for the detection of intracranial injury after head trauma. A meta-analysis published in 2000 examined the association between skull fracture and brain injury.¹⁵ The authors reviewed 200 studies and identified 20 that fulfilled their inclusion criteria. They reported that the sensitivity of skull fracture in detecting patients with an intracranial lesion ranged from 0.13 to 0.75, and the specificity was 0.91 to 0.995. The authors' findings suggested that the presence of a skull fracture increases the probability of an intracranial lesion 5-fold. The meta-analysis concluded that although a fracture demonstrated on plain film increased the likelihood of an intracranial lesion, its low sensitivity precluded its use in ruling out the diagnosis of an intracranial hemorrhage; it is thus of limited clinical value in risk stratification for brain injury. Consequently, current guidelines do not recommend plain film radiographs in the evaluation of patients with MTBI.¹⁶

EMERGENCY DEPARTMENT EVALUATION OF MILD TRAUMATIC BRAIN INJURY

Noncontrast Head Computed Tomography

Although up to 15% of ED-attended MTBI patients with a GCS score of 15 will have head CT acute intracranial pathology identified, less than 1% of these

patients will require neurosurgical intervention.⁵⁻⁷ Early studies using trauma registries established that up to 15% of MTBI patients with a GCS of 14 or 15 will show an acute intracranial traumatic injury on noncontrast head CT; 1% of these will require neurosurgical intervention. The existing literature does not identify which MTBI patients with intracranial lesions clinically deteriorate, nor does it define the relationship between acute traumatic intracranial lesions and the development of postconcussive symptoms. Most studies on patients with MTBI have focused on identifying patients with acute intracranial lesions on CT; ascertaining clinical criteria for the development of postconcussive symptoms has thus far been elusive.

Several clinical decision rules have been developed in an attempt to determine clinical criteria for CT scanning. Two of these rules, the New Orleans Criteria (NOC)⁵ and the Canadian CT Head Rule (CCHR),¹⁷ have been validated prospectively (see Table 2⁵).

In the CCHR study,¹⁷ 3121 patients were prospectively evaluated for having a neurosurgical lesion. Acute intracranial traumatic lesions deemed unimportant were defined by consensus and included smear subdurals and small traumatic subarachnoid hemorrhages. The authors concluded that head CT in MTBI is indicated only in those patients with 1 of 5 high-risk factors: failure to reach a GCS score of 15 within 2 hours of injury, suspected open skull fracture, sign of basal skull fracture, vomiting more than once, or age greater than 64 years.

In deriving the NOC, Haydel *et al.*⁵ prospectively assessed 1429 patients with a GCS score of 15 in the ED and a history of loss of consciousness or amnesia of the traumatic event. A derivation phase in which patients with intracranial injury predictors ($n = 520$) were identified was followed by a validation phase ($n = 909$). The authors reported that 6.5% of patients

Table 2. Criteria for Determining If CT Is Indicated After Minor Head Injury.

New Orleans Criteria: MTBI with a GCS Score of 15	Canadian CT Head Rule: MTBI with a GCS Score of 13 to 15
Headache	High risk (for neurosurgical intervention)
Vomiting	GCS score less than 15 at 2 hours after injury
Age greater than 60 years	Suspected open or depressed skull fracture
Drug or alcohol intoxication	Any sign of basal skull fracture
Persistent anterograde amnesia (deficits in short-term memory)	More than 1 episode of vomiting
Visible trauma above the clavicle	Age greater than 64 years
Seizure	Medium risk (for brain injury on CT)
	Amnesia before impact greater than 30 minutes
	Dangerous mechanism (eg, pedestrian struck by motor vehicle, occupant ejected from motor vehicle, or fall from a height greater than 3 feet or 5 stairs)

NOTE: CT is needed if the patient meets 1 or more of the listed criteria.

Abbreviations: CT, computed tomography; GCS, Glasgow Coma Scale; MTBI, mild traumatic brain injury.

Table 3. *Predictor Variables for Intracranial Lesions.*

	Smits <i>et al.</i> ¹⁸ [OR (95% CI)]	Ibanez <i>et al.</i> ¹⁹ [OR (95% CI)]	Fabbri <i>et al.</i> ⁵¹ [OR (95% CI)]
Class of evidence	II	II	III
Variable			
GCS score of 14	2 (1–3)	7 (4–14)	19 (14–26)
Neurological deficit	2 (1–3)	7 (2–25)	19 (13–28)
Signs of basilar skull fracture	25 (13–47)	11 (6–23)	10 (6–16)
LOC	2 (1–3)	7 (4–11)	2 (2–3)
Posttraumatic amnesia	1.5 (1–2)	3 (2–5)	8 (6–12)
Headache, mild to moderate	1 (0.8–2)	1 (0.8–2)	–
Headache, severe	–	3 (2–6)	–
Vomiting	3 (2–4)	4 (2–7)	5 (3–8)
Posttraumatic seizure	3 (0.8–10)	2 (0.25–17)	3 (2–5)
Alcohol or drug intoxication	1 (0.6–2)	1 (0.3–3)	–
Anticoagulation	2 (1–5)	4 (3–7)	8 (3–9)
Age ≥ 65 years	–	2 (1–3)	2 (1–3)
Dangerous mechanism	2 (1–4)	–	3 (2–4)

Abbreviations: CI, confidence interval; GCS, Glasgow Coma Scale; LOC, loss of consciousness; OR, odds ratio.

Table 4. *Postconcussive Symptoms.*

Somatic	Cognitive	Affective
Headache	Impaired attention	Anxiety
Sleep disturbances	Difficulty concentrating	Depression
Dizziness/vertigo	Memory problems	Emotional lability
Nausea		
Fatigue		
Photophobia/phonophobia		

exhibited intracranial injury on CT, and 0.4% of these patients needed neurosurgical intervention. The NOC consists of 7 identified predictors of an abnormal CT scan, including headache (any head pain), vomiting, age greater than 60 years, intoxication, deficit in short-term memory (persistent anterograde amnesia), physical evidence of trauma above the clavicle, and seizure. A negative predictive value of 100% (95% confidence interval: 99%–100%) accompanies a clinical absence of all 7 criteria.

The presence of any of the clinical elements in the NOC and the CCHR criteria essentially identified all patients requiring emergent neurosurgical intervention. However, both rules have low specificity; if applied appropriately, use of the Canadian rule would result in a 37% reduction in head CT scans, whereas the New Orleans Rule would reduce CT scans by only 3% to 20%. The specificity difference is attributable to the CCHR's primary outcome measure being a lesion requiring neurosurgical intervention, whereas the NOC's primary outcome measure is any identified acute traumatic intracranial lesion.

Studies that provided external validation of the NOC and CCHR have identified several limitations of these rules.^{18,19} Both decision rules use loss

of consciousness or amnesia as entry criteria, and neither applies to patients on anticoagulants. Thus, neither rule can be reliably applied to all patients with head trauma. In addition, if the primary outcome measure of acute traumatic lesion is used (without subcategorization into significant and nonsignificant lesions as stratified in the CCHR), it becomes clear that specificity is sacrificed for sensitivity.

After the NOC and CCHR studies, several database analyses from other countries were published (see Table 3¹⁷ and references therein). These databases highlighted that neither loss of consciousness nor posttraumatic amnesia were sufficient discriminators for determining who needs neuroimaging. Although identified in the NOC, neither seizure nor mild or moderate headache was found to be a significant univariate predictor of intracranial injury. Alcohol was also not found to be a univariate predictor of intracranial injury in 2 of the studies.^{18,19} However, a dangerous mechanism of injury does emerge as an important factor in deciding who requires neuroimaging.

On the basis of the best available current evidence, the 2008 CDC/ACEP joint practice guideline

Table 5. *Discharge Instruction Recommendations for Patients with Mild Traumatic Brain Injury.*

Instructions should be written at approximately a sixth- to seventh-grade level.
Instructions should be given to the patient and immediate caregiver in both print and verbal form. The layout and type fonts should be appropriate for low-literacy materials.
Patients who develop the following symptoms should be instructed to return to the emergency department for re-evaluation: - Repeated vomiting - Worsening headache - Problems remembering - Confusion - Focal neurological deficit - Abnormal behavior - Increased sleepiness or passing out - Seizures
A list of postconcussive symptoms should be provided to the patient in written and verbal form and be used as a prompt for the patient to seek specialist referral (for a list of postconcussive symptoms, see Table 4).
Specialist referral in traumatic brain injury should be provided at discharge to patients who have suffered a head injury.

on MTBI⁶ makes 2 recommendations regarding neuroimaging in the ED:

- Level A. Noncontrast head CT is indicated in head trauma patients with loss of consciousness or posttraumatic amnesia only if 1 or more of the following is present: headache, vomiting, age greater than 60 years, drug or alcohol intoxication, deficits in short-term memory, physical evidence of trauma above the clavicle, posttraumatic seizure, a GCS score less than 15, focal neurological deficit, and coagulopathy.
- Level B. Noncontrast head CT should be considered in head trauma patients with no loss of consciousness or posttraumatic amnesia if there is a focal neurological deficit, vomiting, severe headache, age greater than or equal to 65 years, physical signs of a basilar skull fracture, a GCS score less than 15, coagulopathy, or a dangerous mechanism of injury, which includes ejection from a motor vehicle, a pedestrian struck, and a fall from a height of more than 3 feet or 5 stairs.

Magnetic Resonance Imaging

There are currently no published studies that evaluate magnetic resonance imaging (MRI) within 24 hours of injury in MTBI patients. Therefore, at this time, no evidence-based recommendations can be made

regarding the utilization of MRI in the acute evaluation of MTBI patients, particularly in comparison with the current neuroimaging standard, noncontrast head CT.

Most studies comparing CT to MRI in patients with TBI do not distinguish or stratify MTBI and more severe TBI in their results. MRI has been shown to be more sensitive than CT for exhibiting small subdural hematomas and cortical contusions in patients with MTBI undergoing MRI within 24 hours of ED discharge.²⁰ However, no significant symptom difference has been established between patients with abnormal and normal MRI scans, nor has any impact of these findings on outcomes been established. Several other studies have demonstrated MRI abnormalities that could be attributed to TBI, including contusions, diffuse axonal lesions, and epidural hematomas.^{21,22} However, the clinical relevance of abnormal MRI scans has not been demonstrated in patients with MTBI, and these studies were not temporally proximate to the time of injury; this limits their relevance to the acute management and disposition time frame.

Recent improvements in MRI technology, including decreased scan times and specialized pulse sequences, have improved the detection of structural and functional abnormalities associated with MTBI.²³ One recent study using diffusion tensor imaging in 10 patients with normal CT presenting with a GCS score of 15 established an association between the severity of postconcussion symptoms and diffusion tensor imaging abnormalities.²⁴ Technological refinements, further investigation, enhanced access, and overcoming logistical challenges will determine the future utility of MRI in the MTBI patient's acute evaluation.

BRAIN-SPECIFIC SERUM BIOMARKERS

Serum biomarkers have the potential to eliminate head CT scan requirements in MTBI patients. The proteins and molecules released after a brain injury diffuse into the cerebrospinal fluid, cross the blood-brain barrier, and can be measured with serum assays. Although several brain-specific neuronal and astrocyte proteins and acute phase reactants have been studied, to date S-100B holds the most promise. Serum concentrations of S-100B rise and fall within hours after MTBI, and this makes early measurement critical.²⁵ At a cutoff of 0.10 µg/L, the sensitivity approaches 100%. One class I study and 7 class II studies have reported sensitivities of 90% to 100% but low specificities of 4% to 65% in predicting acute traumatic lesions on head CT.^{26–33} Because

S-100B is also present in adipose, skin, and cartilage, its use is valid only in the absence of extracranial injury.³⁴ The exact role that biomarkers will take in the future evaluation of MTBI is yet to be determined, but they could conceivably play a significant role in resource utilization and in prehospital transport decision algorithms.^{35,36} On the basis of the best available evidence, the 2008 CDC/ACEP joint practice guideline on MTBI recommends the following:⁶

- Level C: In MTBI patients without significant extracranial injuries and a serum S-100B level less than 0.1 µg/L measured within 4 hours of injury, consideration can be given to not performing a CT. (This test has not yet received Food and Drug Administration approval for clinical use in the United States.)

ISSUES IN EMERGENCY DEPARTMENT DISCHARGE PLANNING FOR PATIENTS WITH MILD TRAUMATIC BRAIN INJURY

Several class II and III studies have supported the discharge of MTBI patients after negative head CT.^{37–40} af Geijerstam and Britton⁴¹ conducted a systematic review of the MTBI literature on head CT and collated the outcomes of more than 62,000 MTBI patients with a GCS score of 15. Eleven cases of deterioration after negative CT were identified, but only 3 of the cases were considered definitive. The authors concluded that patients with MTBI who have negative CT can be safely discharged without concern of neurological deterioration.

In a class I multicenter Swedish study involving 39 hospitals, 1292 patients with MTBI were prospectively evaluated for delayed complications post-discharge. Although more than 6% of the patients had a positive intracranial finding, of those with a negative interpretation (and who were discharged from the ED unless other factors required admission), none had developed a complication requiring hospital admission or surgical intervention at the 3-month follow-up.⁴²

Thus far, no study has had sufficient power to identify specific potentially higher risk MTBI patients requiring prolonged hospital observation, such as patients with intrinsic or pharmacological bleeding disorders, patients with previous neurosurgical procedures (such as a ventriculoperitoneal shunt), and those with preexisting neurological disorders or signs. On the basis of the best available evidence, the 2008 CDC/ACEP joint practice guideline on MTBI makes the following recommendation:⁶

Table 6. *Key Points in Dealing with Mild Head Trauma.*

1. Obtain a careful history focusing on loss of consciousness, amnesia, the use of warfarin, a history of hemophilia, and the anatomic location of the injury.
2. Perform a careful physical and neurological examination, looking for signs of a basilar skull fracture (eg, external canal laceration or hemotympanum) and for cranial nerve deficits, specifically in cranial nerves IV and VI.
3. Consider obtaining a noncontrast head computed tomography scan for all patients with a history of loss of consciousness or posttraumatic amnesia. Although the incidence of finding a neurosurgically correctable lesion in patients with a Glasgow Coma Scale score of 15 is less than 1%, the incidence of finding evidence of brain injury approaches 5% to 10% and may have a significant impact in predicting sequelae.
4. Consider imaging infants less than 12 months old with any history or physical findings of head trauma because historical and clinical findings are unreliable.
5. Consider arranging or referring for postdischarge consultation and magnetic resonance imaging in patients with minor head trauma who present with symptoms consistent with postconcussive syndrome.
6. Provide appropriate guidance to athletes regarding sports activities after a concussion, ensuring that they understand the potential consequences related to second impact syndrome.
7. Advise all mild traumatic brain injury patients of the postconcussive syndrome, its time course, and its overall positive prognosis.
8. Preventive strategies are the cornerstone for reducing the morbidity and mortality associated with traumatic brain injury.

- Level B: Patients with an isolated MTBI who have a negative head CT scan result are at minimal risk for developing an intracranial lesion and therefore may be safely discharged from the ED. (There are inadequate data to include patients with a bleeding disorder, patients who are receiving anticoagulation therapy or antiplatelet therapy, or patients who have had a previous neurosurgical procedure in this population.)

POSTCONCUSSIVE SYMPTOMS

Symptoms such as headache, dizziness, anxiety, and impaired cognition and memory may persist after MTBI (see Table 4).⁴³ The proportion of MTBI cases with postconcussive symptoms (< 3 months in duration) and PCS (> 3 months in duration) varies widely in studies and is found to usually diminish over time. It affects more than of 58% of patients 1 month after the injury⁴⁴ and 15% at 1 year.⁴⁵

Table 7. *Areas in Need of Future Research.*

Identification of the subgroup of patients with a Glasgow Coma Scale score of 15 who do not need computed tomography and who can be safely discharged home
Uniform diagnostic criteria for mild traumatic brain injury, including the role of neuroimaging in distinguishing mild traumatic brain injury from moderate traumatic brain injury.
The role of computed tomography in differentiating mild traumatic brain injury from moderate traumatic brain injury
The role of cognitive testing in the diagnosis of mild traumatic brain injury
Patients at risk for second impact syndrome and prevention recommendations
The role of magnetic resonance imaging and other imaging studies in the diagnostic management of traumatic brain injury patients in the emergency department

This information, however, is often omitted from most MTBI discharge instruction sheets. A systematic review concluded that providing educational information on postconcussive symptoms may reduce long-term complaints, but other studies have failed to confirm this conclusion.^{46,47} The presence of headache, nausea, and dizziness during the acute head injury evaluation can be prognostic of PCS: the presence of all 3 symptoms is associated with a 50% likelihood of having PCS at 6 months, whereas an absence of all 3 purports a 28% possibility.⁴⁸

In addition to being challenging to the patient and his family, PCS is a substantial financial and societal burden. As a result of PCS symptoms, patients miss an average of 4.7 work days, and approximately 20% of patients are unemployed at 1 year.⁴⁹ Whether the basis of these problems is structural or functional remains unclear. Regardless of the cause, the patient's complaints should be recognized as a clinical entity for which treatment options exist and follow-up resources should be ideally provided. On the basis of the best available evidence, the 2008 CDC/ACEP joint practice guideline on MTBI makes the following recommendation:⁶

- Level C: MTBI patients discharged from the ED should be informed about postconcussive symptoms.

DISCHARGE INSTRUCTIONS

MTBI patients discharged from the ED should be provided instructions and ideally go home with a responsible adult. One study of discharged MTBI patients reported that 9% were not given any follow-up

recommendations, and 28% were only told "return to the ED as needed." No study has validated criteria for patients to use to return to the ED, although it is generally accepted that changes in mental status, progressive headache, vomiting, new neurological deficit, and seizure are noteworthy signs of which patients should be cognizant. One prospective study⁵⁰ found that MTBI patients rarely remember their discharge instructions, and this implies that it is best to provide post-MTBI discharge instructions in a written form. Table 5 provides the building blocks for discharge instructions recommended in the 2008 CDC/ACEP MTBI guidelines.

CONCLUSION

The lack of uniformity in the definition of MTBI and the absence of large, well-designed trials make the interpretation of much of the MTBI literature difficult at best. The dilemma that the acute care provider faces on a regular basis is which neurologically intact patients with MTBI require CT, who can be safely sent home, and when they can be sent home. Large databases have validated the NOC, which apply to TBI patients with a GCS of 15 who have experienced trauma-related loss of consciousness or posttraumatic amnesia. Large databases have also revealed that neither loss of consciousness nor posttraumatic amnesia is discriminatory for identifying MTBI patients at risk for harboring a neurosurgical lesion, and this makes clinical decision making challenging. No set of clinical criteria has been established that identifies all patients with a neurosurgically significant lesion. There is a growing body of literature supporting the use of brain-specific biomarkers, and these markers may play an important role in the future in identifying patients requiring neuroimaging. All head-injured patients deserve proper counseling regarding postconcussive symptoms with proper follow-up arranged.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Centers for Disease Control and Prevention. Traumatic brain injury in the United States: emergency department visits, hospitalizations, and deaths, October 2004. Available at: http://www.cdc.gov/ncipc/pub-res/tbi_in_us_04/tbi.ed.htm.

2. Centers for Disease Control and Prevention. TBI report to Congress on mild traumatic brain injury in the United States: steps to prevent a serious public health problem, September 2003. Available at: <http://www.cdc.gov/doc.do/id/0900f3ec8006b2e5>. Accessed.
3. Jager T, Weiss H, Coben J, et al. Traumatic brain injuries evaluated in US emergency departments 1992–1994. *Acad Emerg Med* 2000; 7: 134–140.
4. Rutland-Brown W, Langlois J, Thomas K, Lily Y. Incidence of traumatic brain injury in the United States. *J Head Trauma Rehabil* 2006; 21: 544–548.
5. Haydel MJ, Preston CA, Mills TJ, et al. Indications for computed tomography in patients with minor head injury. *N Engl J Med* 2000; 343: 100–105.
6. Stiell I, Clement C, Rowe B, et al. Comparison of the Canadian CT Head Rule and the New Orleans Criteria in patients with minor head injury. *JAMA* 2005; 294: 1511–1518.
7. Smits M, Dippel D, de Haan G, et al. External validation of the Canadian CT Head Rule and the New Orleans Criteria for CT scanning in patients with minor head injury. *JAMA* 2005; 294: 1519–1525.
8. Alves W, Macciocchi S, Barth J. Postconcussive symptoms after uncomplicated mild head injury. *J Head Trauma Rehabil* 1993; 8: 48–59.
9. Rimel RW, Giordani B, Barth JT, et al. Disability caused by minor head injury. *Neurosurgery* 1981; 9: 221–228.
10. Cantu RC. Guidelines for return to contact sports after a cerebral concussion. *Physician Sportsmed* 1986; 14: 75–83.
11. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; 2: 81–84.
12. Centers for Disease Control and Prevention. *Report to Congress on Mild Traumatic Brain Injury in the United States: Steps To Prevent a Serious Public Health Problem*. Atlanta, GA: Centers for Disease Control and Prevention; 2003; 1–47.
13. American Academy of Pediatrics Committee on Quality Improvement, American Academy of Family Physicians Commission on Clinical Policies and Research. The management of minor closed head injury in children. *Pediatrics* 1999; 104: 1407–1415.
14. Jagoda A, Bruns J, Cantrill S, et al. Clinical policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting. *Ann Emerg Med* 2008.
15. Hofman PAM, Nelemans P, Kemerink GJ, et al. Value of radiological diagnosis of skull fracture in the management of mild head injury: meta-analysis. *J Neurol Neurosurg Psychiatry* 2000; 68: 416–422.
16. Jagoda AS, Cantrill SV, Wears RL, et al. Clinical policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting. *Ann Emerg Med* 2002; 40: 231–249.
17. Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT Head Rule for patients with minor head injury. *Lancet* 2001; 357: 1391–1396.
18. Smits M, Diederik W, Dippel W, et al. Predicting intracranial traumatic findings on computed tomography in patients with minor head injury: the CHIP prediction rule. *Ann Intern Med* 2007; 146: 397–405.
19. Ibanez J, Arikian F, Pedraza S, et al. Reliability of clinical guidelines in the detection of patients at risk following mild head injury; results of a prospective study. *J Neurosurg* 2004; 100: 825–834.
20. Doezeema D, King J, Tandberg D, et al. Magnetic resonance imaging in mild head injury. *Ann Emerg Med* 1991; 20: 1281–1285.
21. Voller B, Benke T, Benedetto K, et al. Neuropsychological, MRI and EEG findings after very mild traumatic brain injury. *Brain Inj* 1999; 13: 821–827.
22. Hughes D, Jackson A, Mason D, et al. Abnormalities on magnetic resonance imaging seen acutely following mild traumatic brain injury. *Neuroradiology* 2004; 46: 550–558.
23. Haacke EM, Xu Y, Cheng YC, et al. Susceptibility weighted imaging (SWI). *Magn Reson Med* 2004; 52: 612–618.
24. Wilde E, McCauley S, Hunter J, et al. Diffusion tensor imaging of acute mild traumatic brain injury in adolescents. *Neurology* 2008; 70: 948–955.
25. Levitt MA, Cook LAS, Simon BC, et al. Biochemical markers of cerebral injury in patients with minor head trauma and ethanol intoxication. *Acad Emerg Med* 1995; 2: 675–680.
26. Bazarian JJ, Beck C, Blyth B, et al. Impact of creatine kinase correction on the predictive value of S-100B after mild traumatic brain injury. *Restor Neurol Neurosci* 2006; 24: 163–172.
27. Biberthaler P, Linsenmeier U, Pfeifer K-J, et al. Serum S-100B concentration provides additional information for the indication of computed tomography in patients after minor head injury. A prospective multicenter study. *Shock* 2006; 25: 446–453.
28. Biberthaler P, Mussack T, Wiedemann E, et al. Rapid identification of high-risk patients after minor head trauma (MHT) by assessment of S-100B: ascertainment of a cut-off level. *Eur J Med Res* 2002; 7: 164–170.
29. Biberthaler P, Mussack T, Wiedemann E, et al. Elevated serum levels of S-100B reflect the extent of brain injury in alcohol intoxicated patients after mild head trauma. *Shock* 2001; 16: 97–101.
30. Ingebrigtsen T, Romner B, Marup-Jensen S, et al. The clinical value of serum S-100 protein measurements in minor head injury: a Scandinavian multicentre study. *Brain Inj* 2000; 14: 1047–1055.
31. Muller K, Townsend W, Biasca N, et al. S100B serum level predicts computed tomography findings after minor head injury. *J Trauma* 2007; 62: 1452–1456.
32. Mussack T, Biberthaler P, Kanz KG, et al. Immediate S-100B and neuron-specific enolase plasma measurements for rapid evaluation of primary brain damage in alcohol-intoxicated, minor head-injured patients. *Shock* 2002; 18: 395–400.
33. Poli-de-Figueiredo LF, Biberthaler P, Filho CS, et al. Measurement of S-100B for risk classification of victims sustaining minor head injury—first pilot study in Brazil. *Clinics* 2006; 61: 41–46.
34. Anderson RE, Hansson LO, Nilsson O, et al. High serum S100B levels for trauma patients without head injuries. *Neurosurgery* 2001; 48: 1255–1258.
35. Buduhan G, McRitchie DI. Missed injuries in patients with multiple trauma. *J Trauma* 2000; 49: 600–605.
36. Rizoli SB, Boulanger BR, McLellan BA, et al. Injuries missed during initial assessment of blunt trauma patients. *Accid Anal Prev* 1994; 26: 681–686.

37. Nagy KK, Joseph KT, Krosner SM, et al. The utility of head computed tomography after minimal head injury. *J Trauma Inj Infect Crit Care* 1999; 46: 268–273.
38. Livingston, DH, Loder, PA, Hunt, CD. Minimal head injury: is admission necessary? *Am Surg* 1991; 57: 14–16.
39. Dunham CM, Coates S, Cooper C. Compelling evidence for discretionary brain computed tomographic imaging in those patients with mild cognitive impairment after blunt trauma. *J Trauma Inj Infect Crit Care* 1996; 41: 679–686.
40. Livingston DH, Lavery RF, Passannante MR, et al. Emergency department discharge of patients with a negative cranial computed tomography scan after minimal head injury. *Ann Surg* 2000; 232: 126–132.
41. af Geijerstam JL, Britton M. Mild head injury: reliability of early head computed tomographic findings in triage for admission. *Emerg Med J* 2005; 22: 103–107.
42. af Geijerstam JL, Oredsson S, Britton M, and the OCTOPUS Study Investigators. Medical outcome after immediate computed tomography or admission for observation in patients with mild head injury: randomised controlled trial. *Br Med J* 2006; 333: 4650571.
43. Jagoda A, Riggio S. Mild traumatic brain injury and the postconcussive syndrome. In: Jagoda A, Riggio S, eds. *Emergency Medicine Clinics of North America*. Philadelphia, PA: WB Saunders; 2000; 355–363.
44. Bazarian JJ, Atabaki S. Predicting postconcussion syndrome after minor traumatic brain injury. *Acad Emerg Med* 2001; 8: 788–795.
45. Rutherford WH, Merrett JD, McDonald JR. Symptoms at one year following concussion from minor head injuries. *Injury* 1979; 10: 225–230.
46. Andersson EE, Emanuelson I, Bjorklund R, et al. Mild traumatic brain injuries: the impact of early intervention on late sequelae. A randomized controlled trial. *Acta Neurochir (Wien)* 2007; 149: 151–160.
47. Ghaffar O, McCullagh S, Ouchterlony D, et al. Randomized treatment trial in mild traumatic brain injury. *J Psychosom Res* 2006; 61: 153–160.
48. Carroll LJ, Cassidy JD, Peloso PM, et al. Prognosis for mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J Rehabil Med* 2004; 43(suppl): 84–105.
49. Dikmen SS, Temkin NR, Machamer JE, et al. Employment following traumatic head injuries. *Arch Neurol* 1994; 51: 177–186.
50. Saunders CE, Cota R, Barton CA. Reliability of home observation for victims of mild closed head injury. *Ann Emerg Med* 1986; 15: 160–163.
51. Fabbri A, Servadei F, Marchesini G, et al. Clinical performance of NICE recommendations versus NCWFNS proposal in patients with mild head injury. *J Neurotrauma* 2005; 22: 1419–1427.

Management of Traumatic Brain Injury: First Link in Chain of Survival

Joseph Minardi, MD, and Todd J. Crocco, MD

Department of Emergency Medicine, West Virginia University, Morgantown, WV

ABSTRACT

Traumatic brain injury is a significant cause of morbidity and mortality. The prehospital care of the patient with a traumatic brain injury is critical to maximizing the chances for a good outcome. Prehospital management of the traumatic brain injury patient is directed toward preventing and limiting secondary brain injury while facilitating rapid transport to an appropriate facility capable of providing definitive neurocritical care. Key points in management include the assessment of oxygenation, blood pressure, and mental status (as measured with the Glasgow Coma Scale) and the pupillary examination. Treatment strategies are directed toward maintaining adequate oxygenation and perfusion and treating herniation. Judicious use of temporary hyperventilation and hypertonic saline are considerations. This review provides the most recent evidence regarding the treatment of traumatic brain injury in the prehospital setting and introduces areas in need of future research. *Mt Sinai J Med* 76:138–144, 2009. © 2009 Mount Sinai School of Medicine

Key Words: clinical management, prehospital, review, traumatic brain injury.

Trauma is the leading cause of death in patients 1 to 45 years old.¹ Traumatic brain injury (TBI) is the leading cause of death and disability from trauma with an annual estimated cost of \$60 billion.² The management of TBI requires supportive measures that must be executed meticulously in order to minimize complications and maximize outcomes. The prehospital treatment of TBI victims is critical

because the period immediately following injury is when cerebral tissue is at high risk for secondary injury related to local cerebral ischemia, disruption of cerebrovascular autoregulation, and cerebral edema.³ Interventions in the prehospital setting must focus on limiting secondary brain injury and optimizing cerebral physiology. Many prehospital practices related to the management of patients with TBI are poorly supported by rigorous scientific evidence, and recommendations are based largely on observational studies or extrapolation of emergency department and in-hospital data.

TBI patients often suffer multiple injuries. Their management is complex, and the entire patient should be treated with advanced trauma life support (ATLS) protocols in mind, including spinal precautions and assessment and treatment of chest, abdominal, vascular, and extremity injuries. This article focuses on the specific management of head injuries, but individual patients require complete treatment incorporating the recommendations outlined here as appropriate in the clinical context.

METHODOLOGY

The Brain Trauma Foundation first published evidence-based clinical guidelines for the prehospital management of TBI in 2000, and this was followed by a revision in 2007.¹ These guidelines used material published up to July 2006. Material in this article is primarily a reflection of those recommendations. Additional references were identified from a Medline search spanning 2006 to August 2008 and using the search terms *traumatic brain injury*, *oxygenation*, *hyperventilation*, *hypertonic saline treatment*, *intubation*, *colloid*, and *mannitol*. Articles were included if the data were primarily gathered in the prehospital environment on human subjects or addressed specific questions relevant to the prehospital care of TBI patients. The evidence for this review was graded as high, moderate, or low on the basis of the quality of the available body of literature on a particular

Address Correspondence to:

Todd J. Crocco

Department of Emergency
Medicine, West Virginia
University, Morgantown, WV
Email: tcrocco@hsc.wvu.edu

topic. Recommendations were then graded as strong or weak on the basis of the quality of supporting research available and the degree of confidence regarding a particular recommendation. Because of the lack of direct research in this area, the evidence is predominately low quality, and all of the following recommendations are graded as weak unless otherwise noted.

The recommendations by the Brain Trauma Foundation were organized into assessment, treatment, decision making, and treatment for cerebral herniation. This article is organized in a similar fashion for the ease of the reader and consistency across recommendations (Table 1).

ASSESSMENT

The Brain Trauma Foundation task force identified 4 areas of importance to the prehospital assessment of patients with TBI: oxygenation, blood pressure, Glasgow Coma Scale (GCS) score, and pupillary examination. The prehospital assessment should include pulse oximetry, frequent and accurate measurements of blood pressure, measurements of the GCS score, and pupillary examinations as these areas appear to have the greatest impact on the care of TBI patients.

Pulse Oximetry

The prehospital assessment of a patient with TBI should include continuous pulse oximetry monitoring. Hypoxia at any point during the course of treatment of patients with severe TBI has been associated with secondary brain injury and increased mortality.⁴ In a multicenter study of 150 patients with severe TBI, Chi *et al.*⁵ demonstrated increased mortality and worsened disability in patients suffering hypoxic insults (pulse oximetry <92%) in the prehospital setting. Hypoxic and nonhypoxic patients

were otherwise similar with respect to injury severity and prehospital care. Despite evidence of worsened outcomes, hypoxia is relatively common in the prehospital setting.⁶ Hypoxia in the field may be related to injury pathology as well as underrecognition and undertreatment by prehospital personnel. Future research is needed to delineate specific contributors to prehospital hypoxia, leading to targeted strategies to reduce this phenomenon.

Blood Pressure

Blood pressure should be assessed in the prehospital setting and measured frequently with the most accurate means available. Hypotension, independent of hypoxia, is another factor leading to secondary brain injury.⁷ Hypotension during any period in the course of treatment of TBI has been shown to worsen outcomes and increase mortality. In a prospective study of TBI patients by Chestnut *et al.*,⁴ hypotension was among the most significant predictors of poor outcome. Hypotensive patients had twice the mortality of nonhypotensive patients. Hypotension is generally accepted as a systolic blood pressure less than 90 mm Hg in adults and as shown in Table 2 for pediatric patients.

The intracranial compartment is a fixed space and is very vulnerable to small increases in pressure due to swelling or expansion of masses (eg, hematoma). Because of this, even minor decreases in the systemic blood pressure can have dramatic negative effects on blood flow to the injured brain (recall that the cerebral perfusion pressure is equal to the mean arterial pressure minus the intracranial pressure). For this reason, it is imperative to rapidly identify and correct hypotension in the TBI patient.

Glasgow Coma Scale

The GCS score should be assessed and frequently retaken in TBI patients. The GCS assessment

Table 1. Key Concepts of Prehospital TBI Management.

Assessment	Decision Making	Airway	Breathing	Circulation
• Pulse oximetry • Blood pressure • GCS • Pupils • ETCO₂ (if intubated)	• Minimize out of hospital time • Transport directly to center with neurosurgical capabilities • Transport pediatric patients to pediatric trauma center or adult trauma center with additional pediatric qualifications	• Maintain pulse oximetry >90% • Intubate if unable to maintain adequate airway, maintain pulse oximetry >90%, or if GCS <9	• Avoid hyperventilation • Maintain ETCO₂ 35–40 • In children, BVM may be equal to intubation • Hyperventilation (ETCO₂ 30–35 mm Hg) option for signs of herniation	• Maintain normotension with isotonic crystalloids • Hypertonic saline is treatment option for hypotensive adults with GCS <9

Table 2. *Definitions of Hypotension in Pediatric Patients.*

Age	Systolic Blood Pressure
0–28 days	<60 mm Hg
1–12 months	<70
1–10 years	<70 + 2 × age (years)
>10 years	<90

involves determining the patient's best (1) eye-opening, (2) verbal, and (3) motor responses. GCS assessment requires interaction with the patient, and its measurement differs in preverbal children as shown in Table 3.

The GCS is a brief but useful tool for assessing neurological dysfunction due to TBI. GCS correlates with TBI severity, computed tomography findings, need for neurosurgical intervention, and ultimate outcome.^{8,9} In a retrospective study examining prehospital airway management, Winchell and Hoyt¹⁰ found higher morbidity and mortality for patients with a field GCS score of 3 versus a GCS score of 4 to 8.¹⁰ Similarly, Bazarian *et al.*¹¹ found that patients suffering a loss of consciousness with a subsequent prehospital GCS score of 14 to 15 were unlikely to require neurosurgical intervention. The prehospital GCS score correlates well with the initial GCS score measured in the emergency department.¹¹ Prehospital personnel do require training in determining the GCS score, but after such training (usually with videotaped scenarios); providers are able to accurately measure the GCS score.¹² Clinically, it is useful to measure the best initial GCS score and then make frequent repeat measurements as a decreasing GCS score suggests worsening intracranial pathology and rising

intracranial pressure and may signify impending cerebral herniation.

The GCS score is affected by any factor that diminishes cerebral function, such as hypotension, hypoxia, hypoglycemia, pharmaceuticals, or ingested substances. With this in mind, the GCS score should be measured after initial stabilization of the ABCs, after correction of other underlying factors, and before the administration of sedatives or paralytics.¹

Pupil Examination

The pupils should be examined in the field to gain information regarding diagnosis and prognosis and to assist in treatment decisions. Normal pupils should be symmetric (<1-mm difference) and react briskly to a bright-light stimulus. A normal light reflex requires intact ocular structures, function of cranial nerves II and III, and the brainstem. Pupil examination is affected by direct ocular trauma as well as preexisting ocular pathology. These factors must be considered in the pupillary examination.

Increased intracranial pressure may lead to uncal herniation, which causes compression of cranial nerve III and thus leads to diminished parasympathetic tone manifesting as a dilated and less reactive pupil. Complete destruction of cranial nerve III leads to a fixed and dilated or blown pupil. Pupils that are bilaterally dilated and nonreactive are predictive of increased mortality¹³ and may be due to direct brainstem injury or other toxic or metabolic effects (eg, hypotension, hypoxia, or ingestions). The pupil examination should be interpreted within the entire clinical context.

An abnormal pupil examination correlates with intracranial pathology and outcome¹³ but lacks adequate sensitivity and specificity to predict

Table 3. *Glasgow Coma Scales.*

	Glasgow Coma Scale	Pediatric Glasgow Coma Scale	Score
Eye opening	Spontaneous	Spontaneous	4
	Speech	Speech	3
	Pain	Pain	2
	None	None	1
Verbal	Oriented	Coos, babbles	5
	Confused	Irritable, cries	4
	Inappropriate	Cries to pain	3
	Incomprehensible	Moans to pain	2
	None	None	1
Motor	Obeys commands	Normal spontaneous movement	6
	Localized pain	Withdraws to touch	5
	Flexor withdrawal	Withdraws to pain	4
	Flexor posturing	Abnormal flexion	3
	Extensor posturing	Abnormal extension	2
	None	None	1

the injury pattern or clinical course. The pupil examination has the most utility when used in combination with other physical findings.

TREATMENT

The Brain Trauma Foundation task force identified the following areas of importance to the prehospital treatment of TBI: airway, ventilation, oxygenation, fluid resuscitation, and cerebral herniation. The available evidence suggests that stabilization and treatment of these factors should improve outcomes in TBI.

Airway, Ventilation, and Oxygenation

Even brief hypoxia has severe consequences in the brain-injured patient and must be avoided and corrected as soon as identified. Supplemental oxygen should be administered to maintain pulse oximetry greater than 90%. Any adult patient who has severe TBI (GCS score <9), who is unable to maintain an adequate airway, or who is unable to maintain pulse oximetry greater than 90% despite supplemental oxygen should have an airway established by the most appropriate means as dictated by local emergency medical services (EMS) protocols, training, experience, and available resources.¹

There is significant controversy regarding the role of intubation and the use of rapid sequence induction (RSI) medications (ie, paralytics) in prehospital TBI management. In a series of articles by Davis *et al.*, TBI patients who underwent RSI in the field had increased mortality compared to historical controls,¹⁴ although the use of RSI improved the success rates of intubation.¹⁵ These results are confounded by high rates of hyperventilation and longer scene times in the RSI group. Furthermore, this study was conducted in an urban area and only on ground-transported patients, and this limits its application to rural areas and patients transported by air medical crews. Other studies have contradicted these results, finding either no difference or improved outcomes in TBI patients undergoing RSI.^{16,17} In pediatric TBI, Gaushe *et al.*¹⁸ found no difference in neurological outcome in patients managed with either intubation or bag-valve-mask ventilation. This study was also conducted in an urban practice with relatively short out-of-hospital times but suggests that intubation confers no additional benefit over bag-valve-mask ventilation in certain scenarios. Many factors likely contribute to the different findings in prehospital airway management, such as heterogeneity of provider training and experience

as well as total out-of-hospital time. Specific aspects of intubation and ventilatory management (ie, patient selection, sedatives, and postintubation management) are other elements that may influence outcomes. Future research needs to be done to determine under what conditions, if any, TBI patients benefit from field intubation. Currently, intubation cannot be recommended for ground-transported TBI patients in urban settings who are spontaneously breathing and maintaining pulse oximetry greater than 90%.¹

It seems clear that EMS systems employing RSI and intubation should institute rigorous and ongoing provider training with close monitoring and regular intubation experience.¹⁹

When intubation is performed, both primary and secondary means should be used to confirm placement. Primary confirmation is witnessing tube passage through the glottic opening. Secondary measures include the observation of fog in the endotracheal tube, symmetric chest rise, auscultation of breath sounds, and end-tidal carbon dioxide (ETCO₂) detection. ETCO₂ detectors are more accurate than the means mentioned previously and should be used in addition to physical findings.

Hyperventilation has been shown to worsen outcome in TBI and must be avoided. In a prospective observational study of intubated TBI patients, Warner *et al.*²⁰ found significantly increased mortality in patients with a lower partial pressure of carbon dioxide on emergency department arrival. Despite this, hyperventilation is relatively common in prehospital treatment. Quantitative wave-form capnography correlates well with the partial pressure of carbon dioxide in most patients, is noninvasive, is easily interpreted, and should ideally be used to assess ventilation for intubated patients in the prehospital setting. Goal ventilation should maintain ETCO₂ of 35 to 40 mm Hg unless there are signs of cerebral herniation (discussed later). Many variables affect ETCO₂, including the respiratory rate, tidal volume, pressure, and patient pulmonary status.

ETCO₂ monitoring is not currently the standard in prehospital care (likely because of issues of cost, training, etc.). However, ETCO₂ monitoring provides an objective and easily measured assessment of ventilation, and its widespread implementation is recommended as it should decrease the rates of inadvertent hyperventilation and improve TBI outcomes.

Fluid Resuscitation

Hypotension in the field should be treated with isotonic crystalloids. The cause of hypotension in trauma patients is nearly always blood and fluid

losses, and this makes fluid replacement the most effective way to restore circulation. Normal saline and Ringer's lactate are the most commonly used solutions in the field to treat hypotension.

Colloids have been studied little in the prehospital setting. A 2007 Cochrane review found no evidence to support their use for resuscitation, and a post hoc analysis of the Saline Versus Albumin Fluid Evaluation trial (comparing albumin to crystalloid for resuscitation in intensive care unit patients) found worse outcomes in TBI patients resuscitated with albumin.^{21,22}

Hypertonic saline is a treatment option for hypotensive adults with severe TBI (GCS score <9). In a series of articles by Vassar *et al.*,^{23–26} 7.5% saline was given in a double-blinded fashion as the initial resuscitation fluid for hypotensive TBI patients. The patients who received hypertonic saline had decreased mortality in comparison with the patients receiving normal saline. Results were similar in the hypertonic saline groups with or without the addition of dextran. A larger randomized, double-blinded, placebo-controlled study of hypotensive patients with severe TBI by Cooper *et al.*,²⁷ however, found no benefit to initial resuscitation with 7% saline. These mixed results limit generalized conclusions. Prehospital resuscitation with hypertonic saline has not been studied in pediatric TBI patients and cannot be recommended.

Cerebral Herniation

Cerebral herniation occurs when intracranial pressure becomes severely elevated, exceeding compensatory mechanisms and forcing cerebral tissue downward within the skull. Clinical signs of cerebral herniation include progressive decline in mental status (a decline of more than 2 GCS points from a prior best GCS score of less than 9) dilated or nonreactive pupils, asymmetric pupils, and extensor posturing. For patients with signs of cerebral herniation, more aggressive measures may be warranted. Hyperventilation decreases intracranial pressure by causing cerebral vasoconstriction with the adverse side effect of decreased cerebral blood flow. Although rigorous evidence is lacking, mild hyperventilation (goal $\text{ETCO}_2 = 30\text{--}35$ mm Hg) may be used as a temporizing measure in cases of cerebral herniation. Capnography should be used to monitor ETCO_2 , and periods of hyperventilation should be as brief as feasible. Prophylactic hyperventilation is not recommended and probably worsens outcomes Table 4.

Treatment of cerebral herniation with hyperosmolar therapies (ie, mannitol and hypertonic saline)

Table 4. Respiratory Rates To Induce Hyperventilation.

Age	Respiratory Rate
Adult	20
Child	25
Infant	30

have not been adequately studied in the field. However, inpatient data suggests that mannitol dosed at 1 g/kg as a bolus decreases intracranial pressure (ICP) and may lead to improved functional outcome.²⁸ Hypertonic saline in concentrations from 3% to 23.4% has also decreased intracranial pressure in intensive care unit studies and could be considered to treat cerebral herniation in the emergency department or field.²⁹

EMERGENCY MEDICAL SERVICES DECISION MAKING

Decisions made by EMS dispatchers and providers in the prehospital setting all affect patient care and ultimate outcome. Optimal care can be delivered in organized trauma systems that take information regarding possible brain injury, dispatch the appropriate level of care, and make informed decisions regarding destination. Once on scene, prehospital personnel must make decisions on immediate treatment, transport modality, and destination. Although the evidence is limited, outcomes appear to be improved for patients with shorter prehospital times. A prospective observational study by Härtl *et al.*³⁰ found decreased mortality for severe TBI patients transported directly to a trauma center rather than initially to a nontrauma center that may have been closer. Mortality has been shown to be lower in severe TBI patients transported directly to trauma centers with advanced neurosurgical capabilities, including computed tomography scanners, prompt surgical care, and the ability to monitor intracranial pressure, even in cases in which such a facility is not the nearest. A retrospective review by Potoka *et al.*³¹ found that children with severe TBI had better outcomes when treated in a pediatric trauma center or adult trauma center with additional pediatric qualifications.

CONCLUSION

TBI causes significant morbidity and mortality in the United States. Although care delivered in the prehospital setting has a significant impact on the ultimate outcome of these patients, there are still many

questions to be answered about prehospital TBI management. Current evidence supports meticulous attention to basic (but specifically aimed) supportive care in the field. Prehospital providers need to continue focused efforts to prevent secondary brain injury from hypoxia and hypotension and perform frequent patient assessments including GCS and pupil examinations, all while facilitating rapid transport to the most appropriate facility. EMS providers may also consider treatment with hypertonic saline or hyper-ventilation in certain scenarios. Future study will no doubt change the course of prehospital TBI management, with current therapies moving from the emergency department and hospital to the field and newer therapies improving outcomes in TBI.

ACKNOWLEDGMENT

The authors gratefully acknowledge Stephen M. Davis for his help.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Brain Trauma Foundation. Guidelines for prehospital management of severe traumatic brain injury. *Prehosp Emerg Care* 2007; 12(suppl): S1–S51.
2. National Center for Injury Prevention and Control. Traumatic brain injury. Available at: <http://www.cdc.gov/ncipc/factsheets/tbi.htm#cost>. Accessed August 2008.
3. Moppett IK. Traumatic brain injury: assessment, resuscitation and early management. *Br J Anaesth* 2007; 99: 18–31.
4. Chesnut RM, Marshall LF, Klauber MR, et al. The role of secondary brain injury in determining outcome from severe head injury. *J Trauma* 1993; 34: 216–222.
5. Chi JH, Knudson MM, Vassar MJ, et al. Prehospital hypoxia affects outcome. *J Trauma* 2006; 61: 1134–1141.
6. Dunford JV, Davis DP, Ochs M, et al. Incidence of transient hypoxia and pulse rate reactivity during paramedic rapid sequence intubation. *Ann Emerg Med* 2003; 42: 721–728.
7. Barton CW, Hemphill JC, Morabito D, Manley G. A novel method of evaluating the impact of secondary brain insults on functional outcomes in traumatic brain-injured patients. *Acad Emerg Med* 2005; 12: 1–6.
8. Baxt WG, Moody P. The impact of advanced prehospital emergency care on the mortality of severely brain-injured patients. *J Trauma* 1987; 27: 365–369.
9. Servadei F, Nasi MT, Cremonini AM, et al. Importance of a reliable admission Glasgow Coma Scale score for determining the need for evacuation of posttraumatic subdural hematomas: a prospective study of 65 patients. *J Trauma* 1998; 44: 868–873.
10. Winchell RJ, Hoyt DB. Endotracheal intubation in the field improves survival in patients with severe head injury. Trauma Research and Education Foundation of San Diego. *Arch Surg* 1997; 132: 592–597.
11. Bazarian JJ, Eirich MA, Salhanick SD. The relationship between pre-hospital and emergency department Glasgow Coma Scale scores. *Brain Inj* 2003; 17: 553–560.
12. Lane PL, Báez AA, Brabson T, et al. Effectiveness of a Glasgow Coma Scale instructional video for EMS providers. *Prehospital Disaster Med* 2002; 17: 142–146.
13. Jiang JY, Gao GY, Li WP, et al. Early indicators of prognosis in 846 cases of severe traumatic brain injury. *J Neurotrauma* 2002; 19: 869–874.
14. Davis DP, Hoyt DB, Ochs M, et al. The effect of paramedic rapid sequence intubation on outcome in patients with severe traumatic brain injury. *J Trauma* 2003; 54: 444–453.
15. Davis DP, Ochs M, Hoyt DB, et al. Paramedic-administered neuromuscular blockade improves prehospital intubation success in severely head-injured patients. *J Trauma* 2003; 55: 713–719.
16. Silverboard H, Aisiku I, Martin GS, et al. The role of acute blood transfusion in the development of acute respiratory distress syndrome in patients with severe trauma. *J Trauma* 2005; 59: 717–723.
17. Sloane C, Vilke GM, Chan TC, et al. Rapid sequence intubation in the field versus hospital in trauma patients. *J Emerg Med* 2000; 19: 259–264.
18. Gausche M, Lewis RJ, Stratton SJ, et al. Effect of out-of-hospital pediatric endotracheal intubation on survival and neurological outcome: a controlled clinical trial [published correction appears in *JAMA* 2000;283:3204]. *JAMA* 2000; 283: 783–790.
19. Fakhry SM, Scanlon JM, Robinson L, et al. Prehospital rapid sequence intubation for head trauma: conditions for a successful program. *J Trauma* 2006; 60: 997–1001.
20. Warner KJ, Cuschieri J, Copass MK, et al. The impact of prehospital ventilation on outcome after severe traumatic brain injury. *J Trauma* 2007; 62: 1330–1336; discussion 1336–1338.
21. Perel P, Roberts I. Colloids versus crystalloids for fluid resuscitation in critically ill patients. *Cochrane Database Syst Rev* 2007; 4: CD000567.
22. Myburgh J, Cooper DJ, Finfer S, et al. Saline or albumin for fluid resuscitation in patients with traumatic brain injury. *N Engl J Med* 2007; 357: 874–884.
23. Vassar MJ, Fischer RP, O'Brien PE, et al. A multicenter trial for resuscitation of injured patients with 7.5% sodium chloride. The effect of added dextran 70. The Multicenter Group for the Study of Hypertonic Saline in Trauma Patients. *Arch Surg* 1993; 128: 1003–1011; discussion 1011–1013.
24. Vassar MJ, Perry CA, Holcroft JW. Analysis of potential risks associated with 7.5% sodium chloride resuscitation of traumatic shock [published correction appears in *Arch Surg* 1991;126:43]. *Arch Surg* 1990; 125: 1309–1315.
25. Vassar MJ, Perry CA, Gannaway WL, Holcroft JW. 7.5% sodium chloride/dextran for resuscitation of trauma patients undergoing helicopter transport. *Arch Surg* 1991; 126: 1065–1072.

26. Vassar MJ, Perry CA, Holcroft JW. Prehospital resuscitation of hypotensive trauma patients with 7.5% NaCl versus 7.5% NaCl with added dextran: a controlled trial. *J Trauma* 1993; 34: 622–632; discussion 632–633.
27. Cooper DJ, Myles PS, McDermott FT, et al. Prehospital hypertonic saline resuscitation of patients with hypotension and severe traumatic brain injury: a randomized controlled trial. *JAMA* 2004; 291: 1350–1357.
28. Wakai A, Roberts I, Schierhout G. Mannitol for acute traumatic brain injury. *Cochrane Database Syst Rev* 2007; 1: CD001049.
29. Tyagi R, Donaldson K, Loftus CM, Jallo J. Hypertonic saline: a clinical review. *Neurosurg Rev* 2007; 30: 277–289; discussion 289–290.
30. Härtl R, Gerber LM, Iacono L, et al. Direct transport within an organized state trauma system reduces mortality in patients with severe traumatic brain injury. *J Trauma* 2006; 60: 1250–1256; discussion 1256.
31. Potoka DA, Schall LC, Gardner MJ, et al. Impact of pediatric trauma centers on mortality in a statewide system. *J Trauma* 2000; 49: 237–245.

Neuroimaging of Traumatic Brain Injury

Tuong H. Le, MD, PhD, and Alisa D. Gean, MD

Department of Radiology, Brain and Spinal Cord Injury Center, San Francisco General Hospital, San Francisco, CA

ABSTRACT

In this article, the neuroradiological evaluation of traumatic brain injury is reviewed. Different imaging strategies in the assessment of traumatic brain injury are initially discussed, and this is followed by a review of the imaging characteristics of both primary and secondary brain injuries. Computed tomography remains the modality of choice for the initial assessment of acute head injury because it is fast, widely available, and highly accurate in the detection of skull fractures and acute intracranial hemorrhage. Magnetic resonance imaging is recommended for patients with acute traumatic brain injury when the neurological findings are unexplained by computed tomography. Magnetic resonance imaging is also the modality of choice for the evaluation of subacute or chronic traumatic brain injury. Mild traumatic brain injury continues to be difficult to diagnose with current imaging technology. Advanced magnetic resonance techniques, such as diffusion-weighted imaging, magnetic resonance spectroscopy, and magnetization transfer imaging, can improve the identification of traumatic brain injury, especially in the case of mild traumatic brain injury. Further research is needed for other advanced imaging methods such as magnetic source imaging, single photon emission tomography, and positron emission tomography. *Mt Sinai J Med* 76:145–162, 2009. © 2009 Mount Sinai School of Medicine

Key Words: brain herniation, cerebral edema, diffuse axonal injury, epidural hematoma, intracerebral hematoma, intraventricular hemorrhage, subarachnoid hemorrhage, subdural hematoma, traumatic brain injury.

Traumatic brain injury (TBI) refers to an injury to the intracranial structures following physical trauma to the head. The term *head injury* is preferred when we are addressing injuries that encompass both intracranial and extracranial structures, including the scalp and skull. Clinical classification of the severity of TBI is usually based on the Glasgow Coma Scale (GCS; mild: $12 > \text{GCS} \leq 15$; moderate: $8 > \text{GCS} \leq 12$; severe: $\text{GCS} \leq 8$).¹ In the neuroradiological evaluation of head trauma, TBI is usually classified into primary and secondary injuries (Table 1). Primary injuries are the direct result of trauma to the head. Secondary injuries arise as complications of primary lesions. This classification is important because secondary injuries are potentially preventable, whereas primary injuries, by definition, have already occurred by the time the patient first presents for medical attention. TBI can be further divided according to location (intra-axial or extra-axial) and mechanism (penetrating/open or blunt/closed). The goal of neuroimaging is to identify treatable injuries to prevent secondary damage and to provide useful prognostic information. In this article, a discussion of the different imaging options for TBI is followed by a review of the imaging characteristics of primary and secondary TBI.

Address Correspondence to:

Alisa D. Gean, MD

Department of Radiology
Brain and Spinal Cord Injury
Center

San Francisco General Hospital
San Francisco, CA

Email: alisa.gean@radiology.ucsf.edu

METHODS

A search was performed at <http://www.ncbi.nlm.nih.gov/sites/entrez> (National Center for Biotechnology Information and National Library of Medicine) using 'Text Word' search with the search terms *traumatic* AND *brain*, AND 'Title' search with search term *imaging*, which resulted in 240 articles published in English since 1978. Among these 240 articles,

Table 1. *Imaging Classification of Traumatic Brain Injury*

Primary injury
Extra-axial injury
Epidural hematoma
Subdural hematoma
Subarachnoid hemorrhage
Intraventricular hemorrhage
Intra-axial injury
Axonal injury
Cortical contusion
Intracerebral hematoma
Vascular injury
Dissection
Carotid cavernous fistula
Arteriovenous dural fistula
Pseudoaneurysm
Secondary injury
Acute
Diffuse cerebral swelling/dysautoregulation
Brain herniation
Infarction
Infection
Chronic
Hydrocephalus
Encephalomalacia
Cerebrospinal fluid leak
Leptomeningeal cyst

39 were review articles. Abstracts from the 240 articles were examined. Selected articles from the 240 abstracts and key articles referenced by some of the selected articles, relevant to the objectives of our article, were reviewed. These articles include both pediatric and adult populations.

Most images published in our article were derived from the Teaching File Server (<http://tfserver.ucsf.edu>), which is the property of the Department of Radiology, University of California. Images on the server have been submitted over the years by the attendings, fellows, and residents in the Department of Radiology at Moffit-Long Hospital, the VA Medical Center, and San Francisco General Hospital. Patient demographics were removed in compliance with the Health Insurance Portability and Accountability Act.

IMAGING OPTIONS

Skull Films

Skull films are poor predictors of intracranial pathology and should not be performed to evaluate TBI.²⁻⁴ In mild TBI, skull films rarely demonstrate significant findings. In severe TBI, the lack of abnormality on skull films does not exclude major intracranial injury.⁵ Negative findings may even mislead medical management. Patients who are at

high risk for acute intracranial injury must be imaged by computed tomography (CT).

Computed Tomography

CT is indicated for moderate and severe TBI (GCS ≤ 12) and for patients with mild TBI and age greater than 60 years, persistent neurological deficit, headache or vomiting, amnesia, loss of consciousness longer than 5 minutes, depressed skull fracture, penetrating injury, or bleeding diathesis or anticoagulation therapy.⁶⁻¹² CT is the modality of choice because it is fast, widely available, and highly accurate in the detection of skull fractures and intracranial hemorrhage. Life-support and monitoring equipment can be more easily accommodated in the CT scanner suite than in the magnetic resonance (MR) suite. In addition, CT is superior to magnetic resonance imaging (MRI) in revealing skull fractures and radio-opaque foreign bodies. Intravenous contrast administration should not be performed without a prior noncontrast examination because contrast can both mask and mimic underlying hemorrhage. In suspected vascular injury, CT angiography can be performed at a high (submillimeter) resolution, especially on multidetector CT. However, even with the tremendous advances in CT over the last 3 decades, the majority of mild TBI cases still show no visible abnormality on CT.¹³

Magnetic Resonance Imaging

MRI may be indicated in patients with acute TBI when the neurological findings are unexplained by the CT findings. MRI is also the preferred imaging modality for subacute and chronic TBI. MRI is comparable to CT in the detection of acute epidural hematoma (EDH) and subdural hematoma (SDH).^{14,15} However, MRI is more sensitive to subtle extra-axial smear collections, nonhemorrhagic lesions, brainstem injuries, and subarachnoid hemorrhage (SAH) when fluid attenuated inversion recovery (FLAIR) is used.^{16,17}

Fluid Attenuated Inversion Recovery

FLAIR imaging improves the detection of focal cortical injuries (eg, contusions), white matter shearing injuries, and SAH by suppressing the bright cerebrospinal fluid (CSF) signal typically seen on routine T2-weighted images. Sagittal and coronal FLAIR images are particularly helpful in the detection of diffuse axonal injury (DAI) involving the corpus callosum and the fornix, 2 areas that are difficult to evaluate on routine axial T2-weighted images.

However, an abnormally high signal in the sulci and cisterns of ventilated patients receiving a high inspired oxygen fraction (>0.60) can be observed in uninjured patients and should not be mistaken for hemorrhage.¹⁸

Gradient-Recalled-Echo T2*-Weighted Magnetic Resonance Imaging

Gradient-recalled-echo (GRE) T2*-weighted MRI is highly sensitive to the presence of ferritin and hemosiderin, 2 breakdown products of blood. The presence of hemosiderin and ferritin alters the local magnetic susceptibility of tissue, resulting in areas of signal loss on GRE T2*-weighted images. Because hemosiderin can persist indefinitely, its detection on GRE T2*-weighted images allows for improved evaluation of remote TBI. Unfortunately, GRE images are limited in the evaluation of cortical contusions of the inferior frontal and temporal lobes because of the inhomogeneity artifact induced by the paranasal sinuses and mastoid air cells. This limitation is even more problematic at higher magnetic field strengths unless parallel imaging is used.^{19,20}

Diffusion-Weighted Imaging

Diffusion-weighted imaging (DWI) measures the random motion of water molecules in brain tissue. Because of its improved sensitivity to foci of acute shearing injury, DWI has been particularly useful for the detection of DAI.^{21–24} DWI reveals more DAI lesions than fast spin-echo T2-weighted or GRE T2*-weighted images in patients imaged within 48 hours of injury. The apparent diffusion coefficient, which measures the magnitude of water diffusion averaged over a 3-dimensional space, is often reduced in acute DAI. The fractional anisotropy, which measures the preferential motion of water molecules along the white matter axons, is frequently reduced in chronic DAI. The integrity of white matter tracts can be evaluated with diffusion tensor imaging.

Magnetic Resonance Spectroscopy

Magnetic resonance spectroscopy (MRS) measures the relative amount of metabolites in brain tissue. Common neurochemicals that are measured with proton MRS include *N*-acetylaspartate (NAA), creatinine (Cr), choline, and myoinositol. In brief, NAA is a marker of neuronal health, and Cr is a marker of energy metabolism. A reduction in the NAA/Cr ratio has been found in patients with a history of TBI, and this finding has been correlated with a poorer prognosis.²⁵

Magnetization Transfer Imaging

Magnetization transfer imaging exploits the longitudinal (T1) relaxation coupling between bound (hydration) protons and free water (bulk) protons. When an off-resonance saturation (radiofrequency) pulse is applied, it selectively saturates protons that are bound in macromolecules. These protons subsequently exchange longitudinal magnetization with free water protons. The magnetization transfer ratio provides a quantitative measure of the structural integrity of tissue. A reduction of the magnetization transfer ratio correlates with worse clinical outcome in a patient with a history of TBI.²⁵

Magnetic Source Imaging

Magnetic source imaging (MSI) uses magnetoencephalography to localize weak magnetic signals generated by neuronal electrical activity. In 2 studies, MSI showed excessive abnormal low-frequency magnetic activity in mild TBI patients with postconcussive syndromes.^{26,27} TBI research using MSI has been limited, and the practical application of MSI in the evaluation of TBI has not been very successful.

Single Photon Emission Tomography

Single photon emission tomography (SPECT) is a nuclear medicine study that measures cerebral blood flow (CBF). It can potentially provide a better long-term prognostic predictor in comparison with CT or conventional MRI.²⁸ Specifically, a worse prognosis has been associated with multiple CBF abnormalities, larger CBF defects, and defects that involve the basal ganglia, temporal and parietal lobes, and brainstem. However, SPECT is less sensitive in detecting smaller lesions that are visible on MRI. Therefore, SPECT imaging is complementary for MRI, but not a replacement, in the evaluation of TBI.

Positron Emission Tomography

Positron emission tomography measures regional brain metabolism with 2-fluoro-2-deoxy-d-glucose. In animal studies, acutely injured brain cells show increased glucose metabolism following severe TBI due to intracellular ionic perturbation. Following the initial hyperglycolysis state, injured brain cells show a prolonged period of regional hypometabolism lasting up to months.²⁹ Human studies in TBI have had limited success in demonstrating consistent results regarding regional glucose metabolism. Because of the heterogeneous nature of TBI, studies have found both hypermetabolism and hypometabolism in the same regions across different TBI patients.³⁰ The

metabolic abnormalities can also be found extending far beyond the lesions, especially with SDHs and EDHs.³¹ Cortical contusion, intracerebral hematoma, and encephalomalacia tend to show more regional metabolic abnormalities confined to the specific lesions.

PRIMARY EXTRA-AXIAL INJURY IMAGING FINDINGS

Epidural Hematoma

EDH develops within the potential space located between the inner table of the skull and the dura. The developing hematoma dissects the dura from the inner table of the skull, forming an ovoid mass that displaces the adjacent brain. Because the EDH is located in the potential space between the dura and inner table of the skull, it rarely crosses cranial sutures, where the periosteal layer of the dura is firmly attached at sutural margins (Figure 1). At the vertex, where the periosteum is not tightly attached to the sagittal suture, the EDH can cross the midline. The majority of EDHs are associated with a skull fracture, commonly in the temporal squamosa, where the fracture disrupts the middle meningeal artery.¹² In children, EDHs may occur from stretching or tearing of meningeal arteries without an associated fracture.

On CT, an acute EDH appears as a well-defined, hyperdense, biconvex extra-axial collection (Figures 2 and 3). Mass effect, sulcal effacement, and midline shift are frequently seen with large EDHs. An important imaging finding that predicts rapid expansion of an arterial EDH is the presence of low-density areas within the hyperdense hematoma (the so-called swirl sign), which is thought to represent

active bleeding.^{32,33} It is an ominous sign that needs to be followed closely.

Venous EDHs are less common than arterial EDHs and tend to occur at 3 common locations: the posterior fossa from rupture of the torcula or transverse sinus, the middle cranial fossa from disruption of the sphenoparietal sinus (Figure 3), and the vertex from injury to the superior sagittal sinus.³⁴ Venous EDHs can be difficult to diagnose on axial CT imaging but are readily confirmed on coronal reformatted CT images or multiplanar MR images.

Subdural Hematomas

SDHs usually develop from laceration of bridging cortical veins during sudden head deceleration. They can also arise from injury to pial vessels, pachionian granulations, or penetrating branches of superficial cerebral arteries. Because the inner dural layer and arachnoid are not firmly attached, SDHs are frequently seen layering along the entire hemispheric convexity from the anterior falx to the posterior falx (Figure 4). In elderly patients with cerebral atrophy, the increase in extra-axial space allows for increased motion between the brain parenchyma and the calvarium, resulting in an increased incidence of SDH in these patients. Another cause of SDH is rapid decompression of obstructive hydrocephalus. In this setting, the brain surface recedes from the dura more quickly than the brain parenchyma can re-expand after being compressed by the distended ventricles, and this causes disruption of the bridging cortical veins.

On CT, the acute SDH appears as a hyperdense, homogeneous, crescent-shaped extra-axial collection (Figure 5A). Most SDHs are supratentorial and are located along the convexity. They are also frequently seen along the falx and tentorium (Figure 5B). Because the SDH is often associated with parenchymal injury, the degree of mass effect seen frequently appears more severe with respect to the size of the collection.

In comparison with a normal brain, the density (attenuation) of the acute SDH is higher because of clot retraction. The density of the SDH will progressively decrease as protein degradation occurs. Rebleeding during the evolution of the SDH appears as a heterogeneous mixture of fresh blood and partially liquefied hematoma (Figures 4 and 6). A sediment level or hematocrit effect may be seen from rebleeding or in patients with clotting disorders. The chronic SDH has density similar to, but slightly higher than, that of CSF (Figure 6). The chronic SDH can be difficult to distinguish from prominent

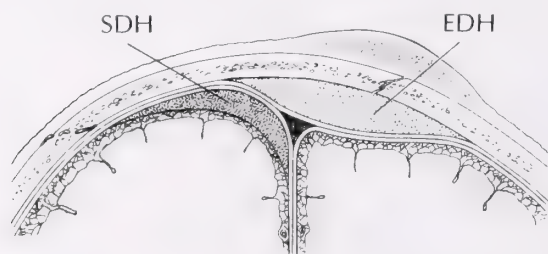


Fig 1. Coronal diagram of EDH and SDH. The EDH is located above the outer dural layer (ie, the periosteum), and the SDH is located beneath the inner (meningeal) dural layer. The EDH does not cross sutures. The SDH does not directly cross the falx or the tentorium.

Abbreviations: EDH, epidural hematoma; SDH, subdural hematoma. Reprinted with permission from Williams & Wilkins-Lippincott.⁴³ Copyright 1994, Williams & Wilkins-Lippincott.

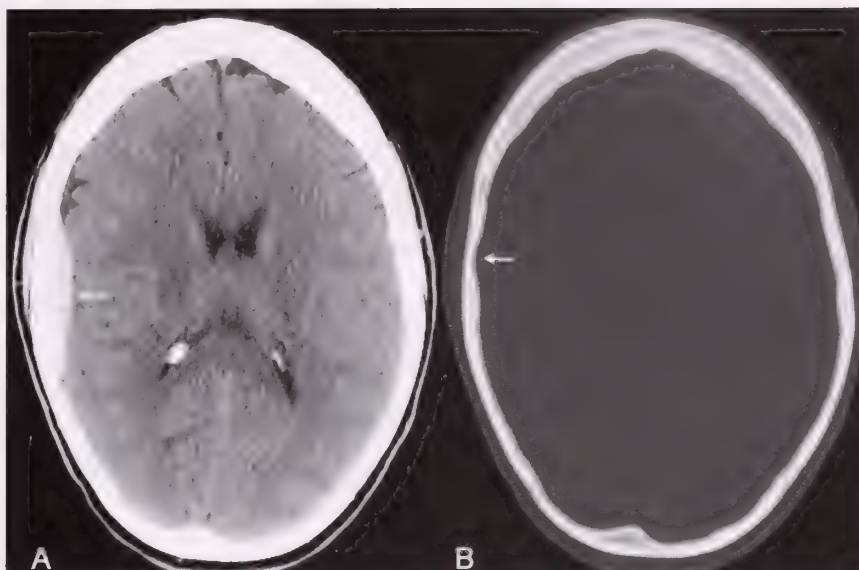


Fig 2. Acute epidural hematoma with an associated linear nondisplaced skull fracture. (A) The axial computed tomography image displayed in the brain window shows a small right temporal epidural hematoma (arrow). (B) The computed tomography scan displayed in the bone window shows a subjacent linear nondisplaced fracture of the temporal squamosa (arrow).

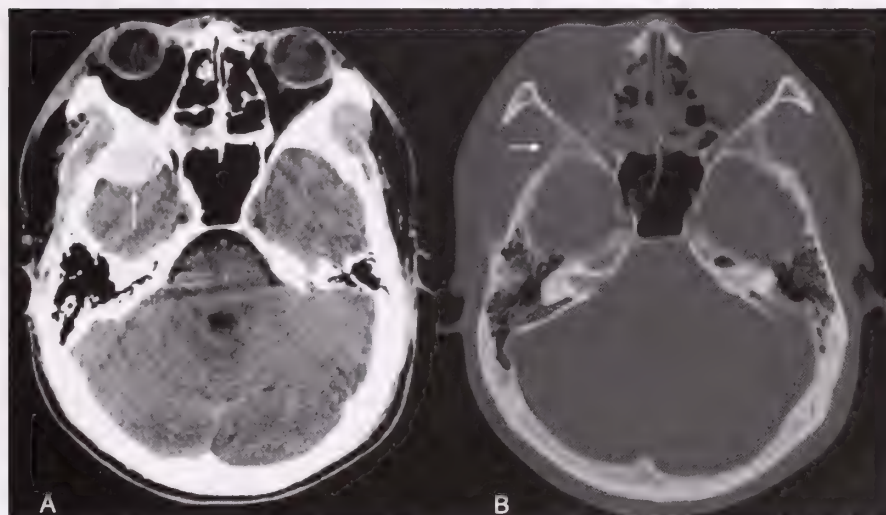


Fig 3. Venous epidural hematoma. (A) The axial computed tomography image shows a biconvex, homogeneous, high-attenuation extra-axial collection within the right middle cranial fossa (arrow). (B) The computed tomography image displayed in the bone window reveals a fracture of the right greater sphenoid wing (arrow). The location suggests that the hematoma is due to disruption of the sphenoparietal sinus.

subarachnoid space in patients with cerebral atrophy. In these patients, contrast-enhanced CT, in addition to noncontrast CT, can improve detection of the chronic SDH by demonstrating an enhancing capsule or displaced cortical veins.

In the subacute setting, usually between 1 and 3 weeks, depending on the patient's hematocrit level, clotting capability, and presence or absence

of rebleeding, an isodense phase occurs. During this transition from the acute SDH to the chronic SDH, a small, thin convexity isodense SDH can be difficult to appreciate (Figure 7). Recognition of indirect imaging findings, such as sulcal effacement, gray matter displacement with white matter buckling, and midline shift, can improve detection of such isodense SDH.

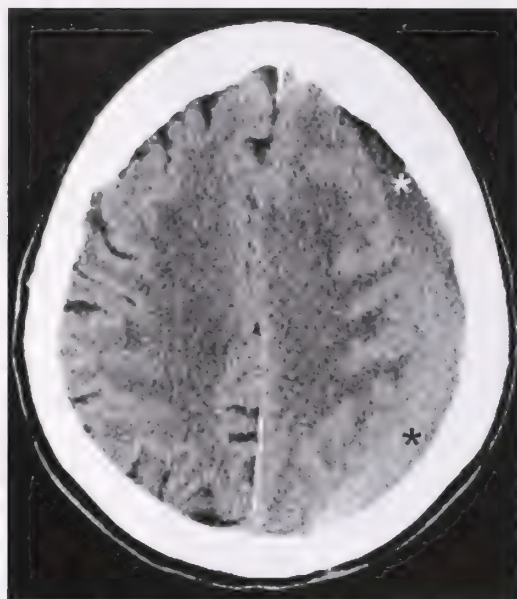


Fig 4. Subdural hematoma with the hematocrit level. The axial computed tomography image shows a large left, holohemispheric extra-axial fluid collection. The collection is a mixture of a dense sediment level (black asterisk) and a partially liquefied hematoma (white asterisk). The attenuation gradient is secondary to the presence of hemorrhage of different ages or to a coagulopathy. There is effacement of the frontoparietal sulci, white matter buckling, and mild left-to-right midline shift resulting from the subdural hematoma mass effect.

On MRI, the acute SDH is isointense to the brain on T1-weighted images and hypointense on T2-weighted images. During the subacute phase, when the SDH may be isodense on noncontrast CT images, the SDH is a high signal intensity on T1-weighted images because of the presence of methemoglobin. The chronic liquefied SDH appears hypointense on T1-weighted images and hyperintense on T2-weighted images with respect to a normal brain (Figure 8). The signal intensity of the chronic SDH is usually slightly higher than the CSF signal intensity on T1-weighted, FLAIR, and proton-density T2-weighted images. Because of its multiplanar capability and its lack of beam-hardening artifact, which limits CT, MRI is useful in identifying small convexity and vertex hematomas that might not be readily detected on axial CT.

Traumatic Subarachnoid Hemorrhage

Traumatic SAH can develop from disruption of small pial vessels, extension into the subarachnoid space by a contusion or hematoma, or transependymal diffusion of intraventricular hemorrhage (IVH). Common sites for SAH include the sylvian and interpeduncular cisterns. The greatest accumulation

of SAH tends to occur contralateral to the site of impact, that is, contrecoup.

On CT, acute SAH appears as linear or serpentine areas of high density that conform to the morphology of the cerebral sulci and cisterns (Figure 6). SAH along the convexity or tentorium can be difficult to differentiate from an SDH. A useful clue is the extension of the SAH into adjacent sulci. Occasionally, effacement of sulci due to the presence of intrasulcal SAH is the only imaging clue to the presence of SAH.

Acute SAH is more difficult to detect on T1-weighted and T2-weighted MR images than on CT because it can be isointense to brain parenchyma. However, FLAIR has been shown to be more sensitive than CT in detecting acute SAH in an animal model, especially when a high volume (1–2 mL) is present.¹⁷ Subacute SAH, when the blood is isointense to CSF on CT, is better appreciated on MRI because of its high signal intensity. Chronic SAH is also much better seen on MRI than on CT. Ferritin and hemosiderin in the subarachnoid space appear as areas of decreased signal intensity on T1-weighted and T2-weighted images (superficial hemosiderosis). As mentioned previously, old blood products are best detected on GRE T2*-weighted images (Figure 9).

Traumatic Intraventricular Hemorrhage

Traumatic IVH can result from rotationally induced tearing of subependymal veins along the surface of the ventricles, by direct extension of a parenchymal hematoma into the ventricular system, or from retrograde flow of SAH into the ventricular system via the fourth ventricular outflow foramina. Besides hydrocephalus, patients with IVH are at risk for developing ependymitis from the irritant effects of the blood.

On CT, IVH usually appears as a CSF-hyperdense fluid level, layering dependently within the ventricular system (Figure 10). Sometimes, a tiny collection of increased density layering in 1 occipital horn may be the only clue to IVH. Occasionally, IVH may appear tumefactive as a cast within the ventricle.

PRIMARY INTRA-AXIAL INJURY

Diffuse Axonal Injury

Axonal injury refers to white matter damage arising from shearing deformation of brain tissue following rotational acceleration and deceleration injury. DAI indicates extensive injury to the white matter and occurs in about half of all severe head trauma cases.³⁵

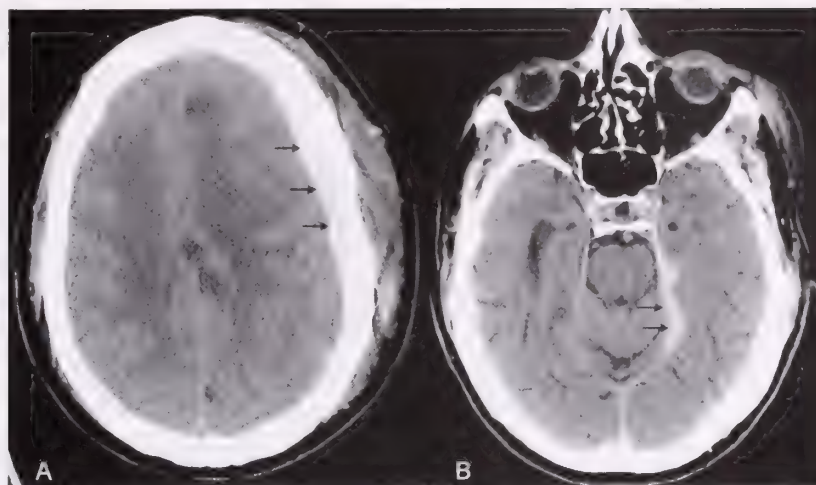


Fig 5. Acute subdural hematoma. (A) The axial computed tomography image demonstrates a left frontoparietal, hyperdense, homogeneous, crescent-shaped extra-axial collection (arrows). An overlying left frontoparietal, subgaleal hematoma is also noted. (B) The axial computed tomography image shows asymmetric hyperdensity along the left tentorium (arrows).

DAI is of special interest because it is considered to be responsible for the majority of TBI cognitive deficits, and yet it is underdiagnosed by conventional imaging techniques.^{36–38}

The location of DAI has been correlated with the severity of the trauma. Specifically, mild DAI (grade I) involves only the peripheral gray-white junctions of the lobar white matter, commonly the parasagittal regions of the frontal lobes and periventricular regions of the temporal lobes. Patients with more moderate DAI (grade II) have abnormalities involving the corpus callosum, particularly the posterior body and splenium, in addition to the lobar white matter. In severe DAI (grade III), the dorsolateral midbrain, in addition to the lobar white matter and corpus callosum, is affected.

On CT, DAI lesions appear as small, petechial hemorrhages and can be seen at the gray-white junction of the cerebral hemispheres (Figure 11), corpus callosum, and dorsolateral midbrain, depending on the severity of the trauma. Because of its higher sensitivity to blood products, GRE T2*-weighted MRI identifies more hemorrhagic DAI lesions than CT.⁴² However, only a minority of DAI lesions are associated with hemorrhage.³⁹ FLAIR MRI can uncover additional nonhemorrhagic foci of DAI but still underestimate the true extent of traumatic white matter damage.^{37,40,41} Nonhemorrhagic acute DAI lesions appear as multiple small foci of increased signal on T2-weighted images and as decreased signal on T1-weighted images. In subacute DAI, intracellular methemoglobin from petechial hemorrhage appears as central hypointensity on T2-weighted images and

as hyperintensity on T1-weighted images. The conspicuity of DAI on MRI eventually diminishes as the damaged axons degenerate and the edema resolves. Chronic DAI imaging findings include nonspecific atrophy, gliosis, and hemosiderin staining, which can persist indefinitely on gradient-echo T2*-weighted images (Figure 11).

Because of its superiority to CT in detecting axonal injuries, MRI helps clinicians explain neurological deficits after trauma and predict long-term outcome. Yet, even with MRI, the incidence of DAI is still underestimated. Newer imaging methods, such as DWI and diffusion tensor imaging with 3-dimensional tractography, have shown potential in improving the detection of white matter injury in both acute and chronic DAI.^{22,23,42} MRS and magnetization transfer imaging can also provide additional prognostic value in DAI.²⁵

Cortical Contusion

Cortical contusion is a focal brain lesion primarily involving superficial gray matter, with relative sparing of the underlying white matter. Regions of the brain that are in close contact with the rough surface on the inner skull table are commonly affected. Therefore, the temporal lobes above the petrous bone or posterior to the greater sphenoid wing are frequently affected. The frontal lobes above the cribriform plate, planum sphenoidale, and lesser sphenoid wing are also commonly involved. Contusions along the parasagittal convexity are less common. Contusions can also occur at the margins of depressed skull fractures. The cerebellum is involved



Fig 6. Chronic subdural hematoma with rebleding, acute subarachnoid hemorrhage, and subgaleal hematoma. The axial computed tomography image shows a small low-density, left frontal extra-axial collection consistent with a chronic subdural hematoma (black arrows). There is also a tiny right frontal subdural collection with mixed high and low densities, which is consistent with rebleding within a chronic subdural hematoma (horizontal white arrows). Linear high attenuation conforming to the sulci of the posterior right frontal lobe is consistent with acute subarachnoid hemorrhage (vertical white arrows). A large subgaleal scalp hematoma is noted at the coup site (asterisk).

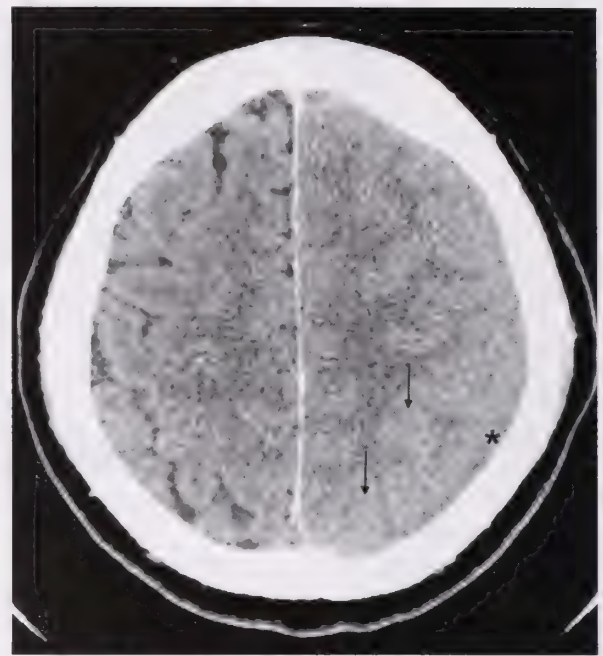


Fig 7. Isodense subacute subdural hematoma. During the transition from an acute subdural hematoma to a chronic subdural hematoma, an isodense phase occurs. At this stage, the subdural hematoma (asterisk) can be difficult to differentiate from the adjacent parenchyma. Note the effacement of the subjacent sulci (arrows), which serves as an important clue in detecting subtle subdural hematomas.

persist indefinitely and serves as an important marker of prior TBI.

less than 10% of the time.⁴³ Contusions are associated with a better prognosis than DAI unless they are accompanied by brainstem injury or significant mass effect.

On CT, hemorrhagic contusions appear as areas of high density within superficial gray matter (Figure 12). They may be surrounded by larger areas of low density from the associated vasogenic edema. As the contusion evolves, the characteristic salt and pepper pattern of mixed areas of hypodensity and hyperdensity becomes more apparent. Nonhemorrhagic contusions appear as low-attenuation areas and can be difficult to detect initially until the development of associated edema.

On MRI, contusions appear as ill-defined areas of variable signal intensity on both T1-weighted and T2-weighted images, depending on the age of the lesions (Figure 12). Again, because contusions are limited to the surface of the brain, they often have a gyral morphology. Hemosiderin from an old contusion can

Intracerebral Hematoma

The intracerebral hematoma is shear-induced hemorrhage due to the rupture of small intraparenchymal blood vessels. Because the bleeding occurs in areas of relatively normal brain, intracerebral hematomas usually have less surrounding edema than cortical contusions. Most traumatic intracerebral hematomas are located in the frontotemporal white matter. Involvement of the basal ganglia has been described; however, hemorrhage in basal ganglia should alert the clinician that an underlying hypertensive bleed may be the culprit (Figure 13). Intracerebral hematomas are often associated with skull fractures and other primary intracranial injuries, including contusions and DAI, especially in patients who are unconscious at the time of injury. Intracerebral hematoma is the most common cause of clinical deterioration in patients who have experienced a lucid interval after the initial injury. Delayed hemorrhage is a common cause of clinical deterioration during the first several days after head trauma.

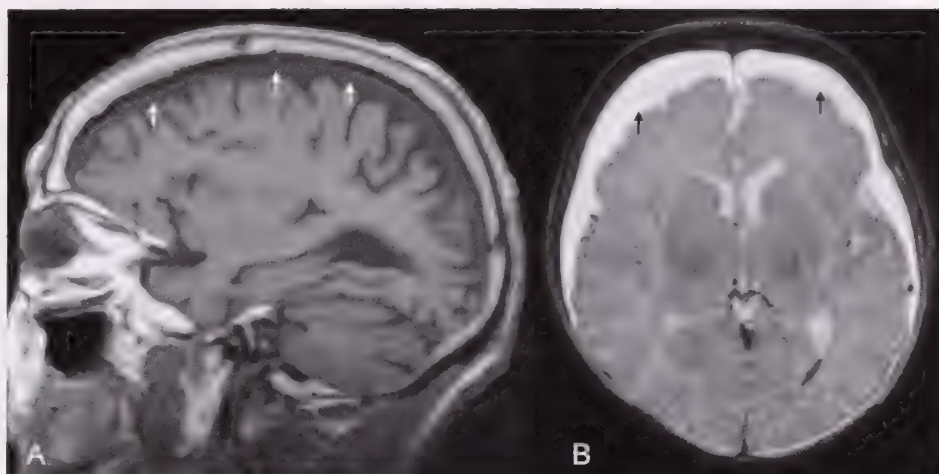


Fig 8. Appearance of the chronic subdural hematoma on magnetic resonance imaging. The chronic subdural hematoma is always slightly higher in signal intensity than cerebrospinal fluid on T1-weighted and T2-weighted images. It is (A) hypointense to gray and white matter on T1-weighted images (arrows) and (B) hyperintense to brain parenchyma on T2-weighted images. Unlike prominent subarachnoid spaces that are characteristic of cerebral atrophy, the chronic subdural hematoma causes effacement of the sulci. In addition, when rebleeding has occurred, septations may be identified within the collection (arrows).

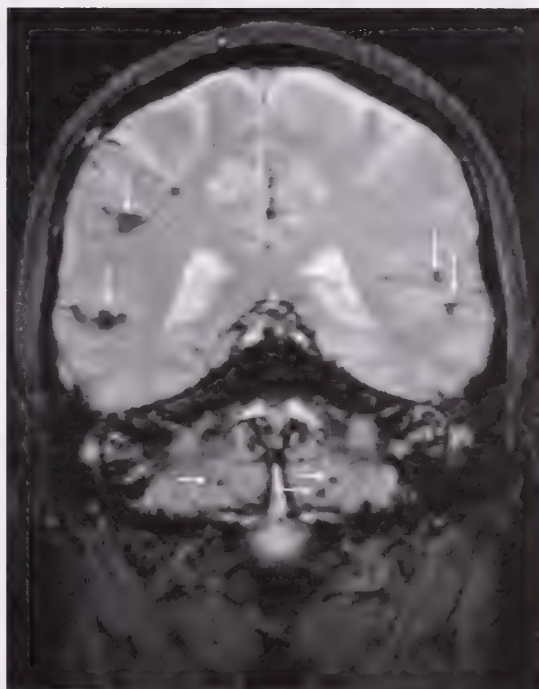


Fig 9. Chronic subarachnoid hemorrhage on magnetic resonance imaging. The coronal gradient-recalled-echo T2*-weighted image shows a low signal (vertical arrows) within the subarachnoid space of the temporal lobes and right frontal lobe due to hemosiderin deposits (superficial hemosiderosis). Foci of ferritin and hemosiderin staining within the cerebellum are also noted (white arrows), with particular involvement of the vermi.



Fig 10. Acute intraventricular hemorrhage. The axial computed tomography image shows hyperdense fluid layering within the occipital horn of the left lateral ventricle (asterisk).

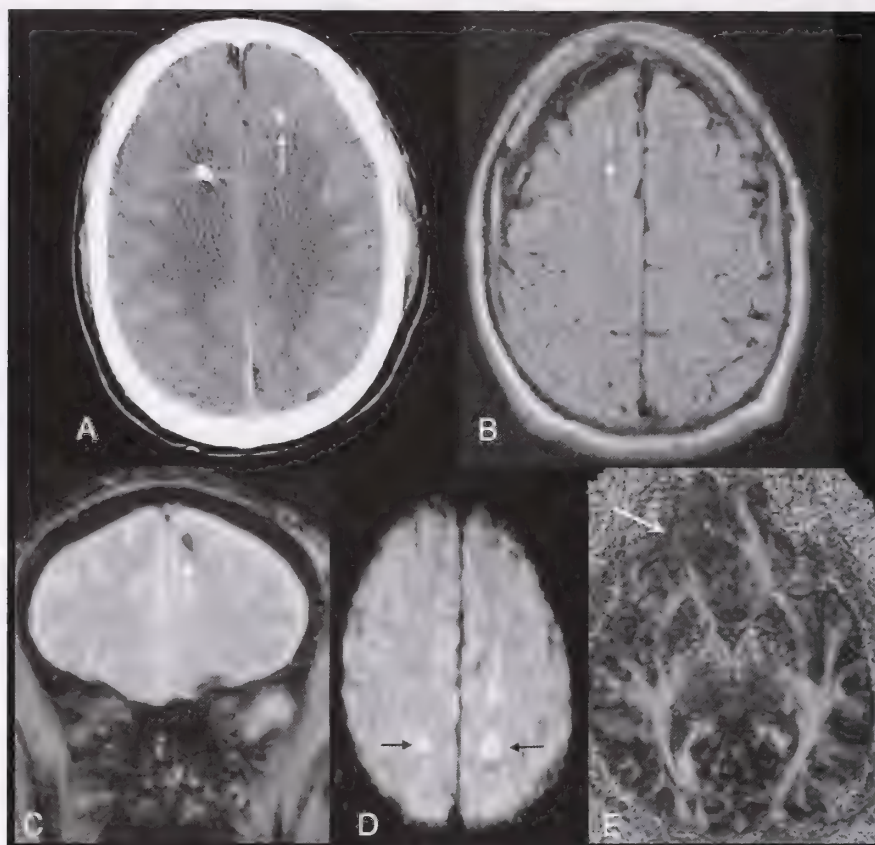


Fig 11. Diffuse axonal injury on computed tomography and magnetic resonance imaging. (A) The axial computed tomography image reveals a small hyperdense focus (arrow) at the gray-white junction of the left frontal lobe consistent with hemorrhagic diffuse axonal injury. Bilateral high attenuation with the sulci of the frontal and parietal lobes represents acute subarachnoid hemorrhage. The intensely dense focus with streak artifact within the right frontal lobe is from a ventricular drainage catheter. (B) The axial fluid attenuated inversion recovery image shows an additional lesion (arrow) within the peripheral gray-white junction of the right frontal lobe. (C) The coronal gradient-recalled-echo T2*-weighted image demonstrates a low signal focus (arrow) at the gray-white junction of the left superior frontal gyrus consistent with deoxyhemoglobin from hemorrhagic diffuse axonal injury. (D) The diffusion-weighted image shows reduced diffusion within the parietal lobe subcortical white matter bilaterally (arrows). (E) In a different patient with a history of moderate traumatic brain injury and diffuse axonal injury, the color fractional anisotropy map shows a reduction in fractional anisotropy in the white matter tract of the right inferior frontal lobe (arrow). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

VASCULAR INJURY

Vascular injuries cause both intra-axial and extra-axial injuries, including the source of hematomas and SAH described previously. Traumatic vascular injuries include arterial dissection, pseudoaneurysm, and arteriovenous fistula. Vascular injuries are usually related to skull base fractures. The internal carotid artery is the most commonly affected vessel. The injury usually occurs at sites of fixation, where the internal carotid artery enters the carotid canal at the base of the petrous bone and at its exit from

the cavernous sinus beneath the anterior clinoid process.

MR findings of traumatic vascular injury include (1) the presence of an intramural hematoma, which is best seen on T1-weighted images with fat suppression (Figure 14), (2) intimal flap with dissection, and (3) the absence of a normal vascular flow void secondary to slow flow or occlusion. An associated parenchymal infarction supplied by the injured vessel may also be seen.

Conventional angiograms are the gold standard for confirmation and delineation of the vascular

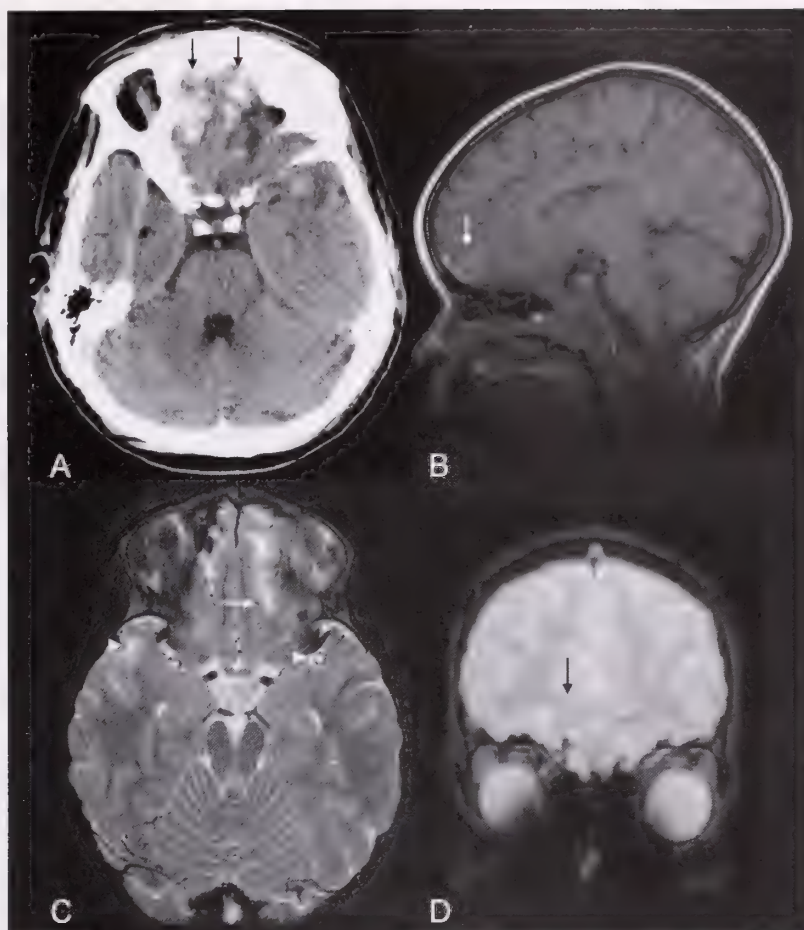


Fig 12. Computed tomography and magnetic resonance imaging of a cortical contusion. (A) The axial computed tomography image demonstrates high attenuation within the gyral crests of the bilateral orbitofrontal lobes (arrows). Partial volume artifact and streak artifact in this area are particularly problematic on computed tomography. (B) The sagittal T1-weighted image demonstrates abnormal increased signal intensity involving the right orbitofrontal lobe (arrow). (C) The axial T2-weighted image demonstrates bilateral orbitofrontal (arrows) and temporal (arrowheads) high signal. (D) The coronal gradient-recalled-echo T2*-weighted image shows decreased signal within the right orbitofrontal lobe that is consistent with blood products (arrow).

dissection and may also show spasm or pseudoaneurysm formation. However, MR angiography and multidetector CT angiography serve as important screening tools in the evaluation of patients with suspected vascular injury.

Traumatic Carotid Cavernous Fistula

The traumatic carotid cavernous fistula is a direct communication between the cavernous portion of the internal carotid artery and the surrounding venous plexus. It typically results from a full-thickness arterial injury, and it may be seen with both blunt and penetrating TBI. The fistula leads to venous engorgement of the cavernous

sinus (Figure 15) and, sometimes, the ipsilateral superior ophthalmic vein and inferior petrosal sinus. Other imaging findings include retrobulbar fat stranding, proptosis, preseptal soft tissue swelling, extraocular muscle enlargement, and asymmetry of the cavernous sinus. Findings may be bilateral because venous channels connect the cavernous sinuses. In severe cases that cause intracranial venous hypertension, brain edema and hemorrhage may be seen. Skull base fractures, especially those involving the sphenoid bone, should alert the clinicians to search for associated cavernous carotid injury. Patients can present with findings weeks or even months after the initial trauma. Therefore, a carotid

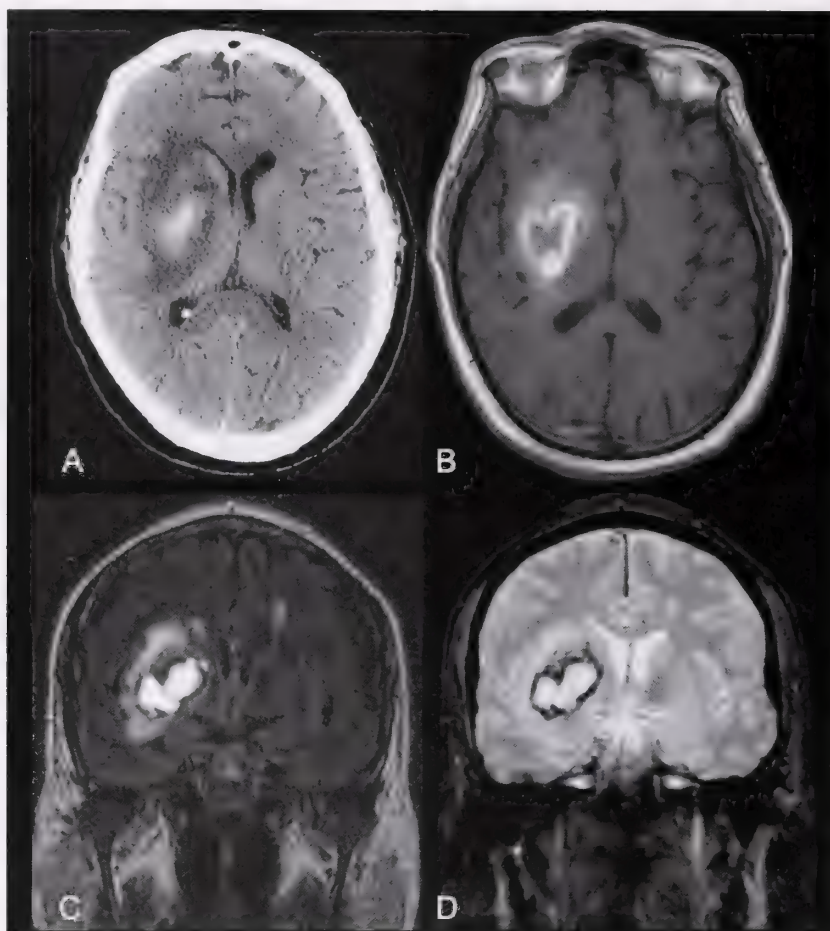


Fig 13. Intracerebral hematoma. (A) The axial computed tomography image shows a hyperdense lesion within the right putamen. The surrounding low attenuation is due to vasogenic edema. (B) On the T1-weighted image, the periphery of the hematoma is hyperintense because of the presence of methemoglobin. (C) The hematoma is also hyperintense on the fluid attenuated inversion recovery image. (D) The gradient-echo T2*-weighted image shows a surrounding hemosiderin rim, which is dark because of local magnetic susceptibility inhomogeneity. Reprinted with permission from *Seminars in Roentgenology*.⁴⁵ Copyright 2006, Elsevier.

cavernous fistula can be overlooked if a detailed clinical history and ophthalmic examination are not performed.

Arteriovenous Dural Fistula

Another traumatic vascular injury is the arteriovenous dural fistula. It is often caused by laceration of the middle meningeal artery with resultant meningeal artery to meningeal vein fistulous communication (Figure 16). Because the fistula generally drains via the meningeal veins, the injured middle meningeal artery rarely leads to the formation of an EDH. Patients are often asymptomatic or present with nonspecific complaints such as tinnitus.

ACUTE SECONDARY INJURY

Diffuse Cerebral Swelling

Diffuse cerebral swelling develops from an increase in cerebral blood volume (hyperemia) or an increase in tissue fluid (vasogenic edema and cytotoxic edema). Effacement of the cerebral sulci and cisterns and compression of the ventricles are typical findings on imaging. Hyperemia and vasogenic edema are thought to be the result of cerebral dysautoregulation, and cytotoxic edema is believed to occur secondary to tissue hypoxia. In cytotoxic edema, the gray-white differentiation is lost; this is in contrast to hyperemia, in which the gray-white differentiation is preserved. Even with cytotoxic edema, the cerebellum and brainstem are usually spared and may appear

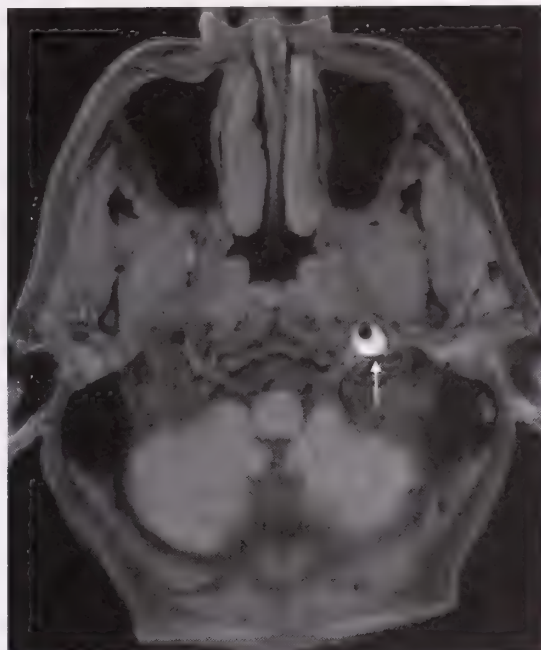


Fig 14. Carotid dissection. The T1-weighted magnetic resonance image, performed with fat suppression, shows a hyperintense crescent (arrow) beneath the adventitia of the left internal carotid artery. The high signal represents a subacute intramural hematoma. In this case, the caliber (flow void) of the vessel lumen is preserved.



Fig 15. Acquired carotid cavernous fistula. Computed tomography angiography demonstrates asymmetric enhancement of the left cavernous sinus secondary to abnormally dilated venous channels (arrow); compare this with the normal appearance of the right cavernous sinus.

hyperintense with respect to the affected cerebral hemispheres (Figure 17).

Traumatic Brain Herniation

Traumatic brain herniation occurs secondary to the mass effect produced by both primary and secondary injuries. In subfalcine herniation, the most common form of herniation, the cingulate gyrus is displaced across the midline under the falx cerebri (Figure 18). Compression of the ipsilateral ventricle due to the mass effect and enlargement of the contralateral ventricle due to obstruction of the foramen of Monro can be seen on imaging. In uncal herniation, the medial temporal lobe is displaced over the free margin of the tentorium. Effacement of the ambient and lateral suprasellar cisterns is an important clue to the presence of uncal herniation. In transtentorial herniation, the brain herniates either upward or downward because of lesions within the posterior fossa or the supratentorium. Upward herniation occurs when portions of the cerebellum and vermis displace through the tentorial incisura. With posterior fossa downward herniation, the cerebellar tonsils displace through the foramen magnum. Downward herniation of the cerebrum manifests as effacement of the suprasellar and perimesencephalic cisterns. Inferior displacement of the calcified pineal gland is

another clue to the presence of downward herniation. With brain herniation, the underlying culprit must be corrected in a timely manner to prevent additional secondary injuries.

Ischemia and Infarction

Ischemia and infarction can occur because of diffuse increased intracranial pressure or a focal compressive mass effect on cerebral vasculature by herniation or hematoma. In subfalcine herniation, the anterior cerebral arteries are displaced to the contralateral side, trapping the callosomarginal branches of the anterior cerebral arteries and leading to anterior cerebral artery infarction. In severe uncal herniation, displacement of the brainstem can compress the contralateral cerebral peduncle and the posterior cerebral artery against the tentorium (Kernohan's notch), leading to peduncular infarction and/or posterior cerebral artery infarction. The third cranial nerve may also be compressed; these patients typically present with a blown pupil and ipsilateral hemiparesis. Tonsillar herniation can cause ischemia in the territory of the posterior inferior cerebellar artery distribution.

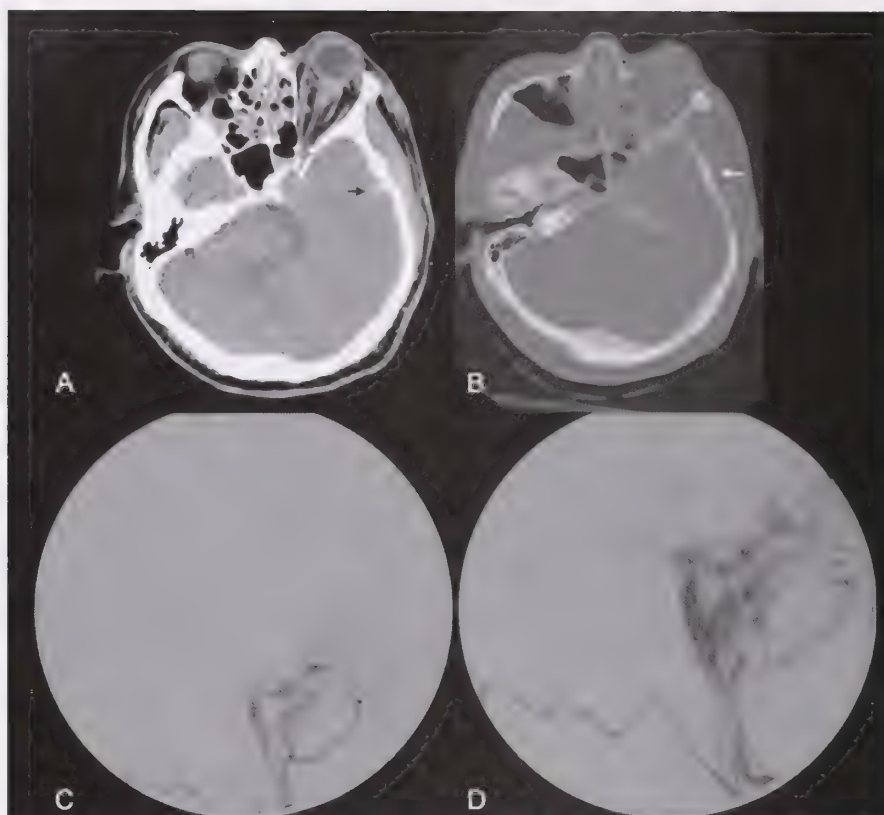


Fig 16. Dural arteriovenous fistula. (A) The axial computed tomography image displayed in the soft tissue window shows a small, round, dense, left temporal extra-axial focus (arrow). Left orbital proptosis and retrobulbar soft tissue stranding can also be seen. (B) The corresponding bone window image shows an associated temporal bone fracture (arrow). (C,D) Images from an external carotid artery cerebral angiogram in the lateral projection show blush of contrast due to filling of the middle meningeal vein via the middle meningeal artery.

CHRONIC SECONDARY INJURY

Traumatic Hydrocephalus

Traumatic hydrocephalus occurs secondary to impaired CSF reabsorption at the level of the arachnoid villi (communicating hydrocephalus) or secondary to obstruction of the cerebral aqueduct and fourth ventricular outflow (noncommunicating hydrocephalus). Hydrocephalus is not an uncommon complication seen in patients with prior SAH or IVH. A mass effect from brain herniation or a hematoma can also cause noncommunicating hydrocephalus via compression of the aqueduct and ventricular outflow foramina. On imaging, the ventricles are dilated, with associated effacement of the sulci in severe cases (Figure 19).

Encephalomalacia

Encephalomalacia is a common but nonspecific sequela of prior parenchyma injury. It is often

clinically asymptomatic, but it can also be a potential seizure focus. The CT imaging appearance of encephalomalacia consists of a relatively well-defined area of low attenuation with volume loss. Encephalomalacia follows CSF signal intensity on both CT and MR, except for the areas of gliosis, which appear as low intensity on T1-weighted images and as high intensity on T2-weighted images. Encephalomalacia within the orbitofrontal (especially the gyrus rectus) and anteroinferior temporal lobes is characteristic of remote traumatic injury.

Cerebrospinal Fluid Leak

A CSF leak results from a dural tear secondary to a skull base fracture. CSF otorrhea occurs when communication between the subarachnoid space and middle ear develops in association with a ruptured tympanic membrane. CSF rhinorrhea arises when there is communication between the subarachnoid space and the paranasal sinuses. In patients with recurrent meningeal infections and a prior history

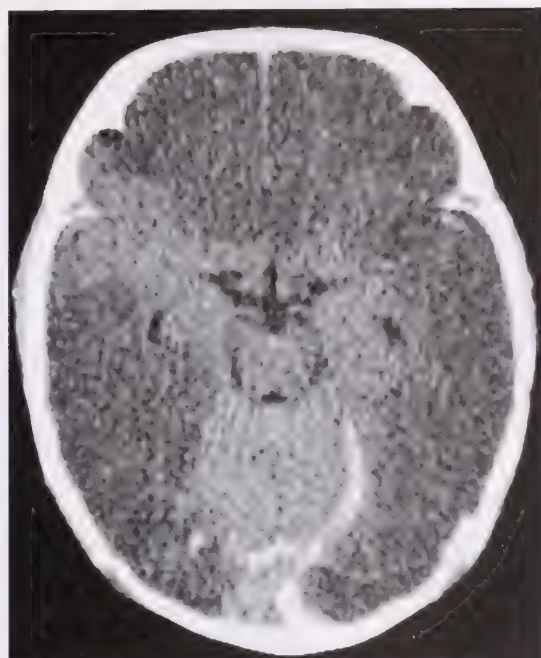


Fig 17. Diffuse cerebral edema. The axial computed tomography image of an infant with a history of strangulation shows a diffuse decrease in attenuation of the cerebral hemispheres with loss of gray-white differentiation. Sparing of the bilateral uncus, right frontal-temporal junction, brainstem, and cerebellum causes these structures to appear dense with respect to the rest of the brain. An acute subdural hematoma along the left tentorium is also noted.

of trauma, a CSF leak should be suspected. CSF leaks are often difficult to localize. Radionuclide cisternography is highly sensitive for the detection of CSF leaks. However, CT scanning with intrathecal contrast is required for detailed anatomic localization of the defect (Figure 20).

Leptomeningeal Cyst

The leptomeningeal cyst, a pediatric lytic calvarial lesion commonly called a growing fracture, is briefly mentioned here because it is also caused by a tear in the dura. The dural defect allows expansion of the arachnoid at the site of the bony defect, presumably as a result of CSF pulsation. Such expansion leads to progressive, slow widening of the skull defect or suture. The leptomeningeal cyst appears as a lytic skull defect on CT or plain skull films.

CONCLUSION

In this article, imaging options for TBI (Table 2) and classification of TBI have been presented. In the management of acute TBI, the goal of imaging is to identify treatable injuries to prevent secondary damages. Despite the advances in MR technology over the last 2 decades, CT continues to be the modality of choice in the evaluation of acute injury because it is fast, widely available, and highly sensitive to acute blood and can more easily accommodate life-support and monitoring equipment. MRI is indicated for patients with acute TBI when the neurological findings are unexplained by CT. MRI is also the modality of choice for subacute or chronic injury. Mild TBI continues to be underdiagnosed by imaging. Advances in MR methods, such as DWI, MRS, and magnetization transfer imaging, can further improve the neuroradiological evaluation of TBI and enhance our understanding of the pathophysiological manifestations of brain trauma. SPECT and positron emission tomography imaging can provide additional

Table 2. *Imaging Options for Traumatic Brain Injury.*

Skull film
Not recommended
Computed tomography
Recommended in the acute setting
Computed tomography cisternography for suspected cerebrospinal fluid leak
Computed tomography angiogram for vascular injury
Preferred examination for skull fractures
Magnetic resonance imaging
Recommended in the acute setting when neurological findings are unexplained by computed tomography
Recommended for subacute and chronic traumatic brain injury
Recommended for nonaccidental trauma
T1-weighted with fat suppression for suspected vascular dissection
Magnetic resonance angiogram for suspected vascular injury
Diffusion-weighted imaging for posttraumatic infarction and axonal injury
Magnetic resonance spectroscopy and magnetization transfer imaging
May be helpful in diffuse axonal injury in predicting long-term prognosis
Single photon emission tomography, positron emission tomography, and magnetic source imaging
Usefulness still limited



Fig 18. Subfalcine herniation due to a subdural hematoma. The axial computed tomography images show an acute, left, heterogeneous, holohemispheric subdural hematoma with associated sulcal effacement, midline shift, and left-to-right subfalcine herniation (arrows).

prognostic value, although additional research in these areas is still needed.

ACKNOWLEDGMENT

We thank the residents, fellows, and attendings from the Neuroradiology Section of the Department of Radiology, University of California San Francisco, for their continuing effort in submitting interesting cases to the teaching file server (<http://tfserver.ucsf.edu>). Some of the cases presented in this article were the products of their effort.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; 2: 81–84.
2. Bell RS, Loop JW. The utility and futility of radiographic skull examination for trauma. *N Engl J Med* 1971; 284: 236–239.
3. Hackney DB. Skull radiography in the evaluation of acute head trauma: a survey of current practice. *Radiology* 1991; 181: 711–714.
4. Masters SJ. Evaluation of head trauma: efficacy of skull films. *AJR Am J Roentgenol* 1980; 135: 539–547.
5. Adams J. Pathology of nonmissile head injury. *Neuroimaging Clin N Am* 1991; 1: 397–410.
6. Haydel MJ, Preston CA, Mills TJ, et al. Indications for computed tomography in patients with minor head injury. *N Engl J Med* 2000; 343: 100–105.
7. Jagoda AS, Cantrill SV, Wears RL, et al. Clinical policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting. *Ann Emerg Med* 2002; 40: 231–249.
8. Lorberboym M, Lampl Y, Gerzon I, Sadeh M. Brain SPECT evaluation of amnesic ED patients after mild head trauma. *Am J Emerg Med* 2002; 20: 310–313.
9. Stiell IG, Lesiuk H, Wells GA, et al. Canadian CT head rule study for patients with minor head injury: methodology for phase II (validation and economic analysis). *Ann Emerg Med* 2001; 38: 317–322.
10. Stiell IG, Lesiuk H, Wells GA, et al. The Canadian CT Head Rule Study for patients with minor head injury: rationale, objectives, and methodology for phase I (derivation). *Ann Emerg Med* 2001; 38: 160–169.
11. Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT Head Rule for patients with minor head injury. *Lancet* 2001; 357: 1391–1396.
12. Zee CS, Go JL. CT of head trauma. *Neuroimaging Clin N Am* 1998; 8: 525–539.
13. Lee H, Wintermark M, Gean AD, et al. Focal lesions in acute mild traumatic brain injury and neurocognitive outcome: CT versus 3T MRI. *J Neurotrauma* 2008; 25: 1049–1056.
14. Gentry LR, Godersky JC, Thompson B, Dunn VD. Prospective comparative study of intermediate-field MR and CT in the evaluation of closed head trauma. *AJR Am J Roentgenol* 1988; 150: 673–682.
15. Orrison WW, Gentry LR, Stimac GK, et al. Blinded comparison of cranial CT and MR in closed head injury evaluation. *AJNR Am J Neuroradiol* 1994; 15: 351–356.

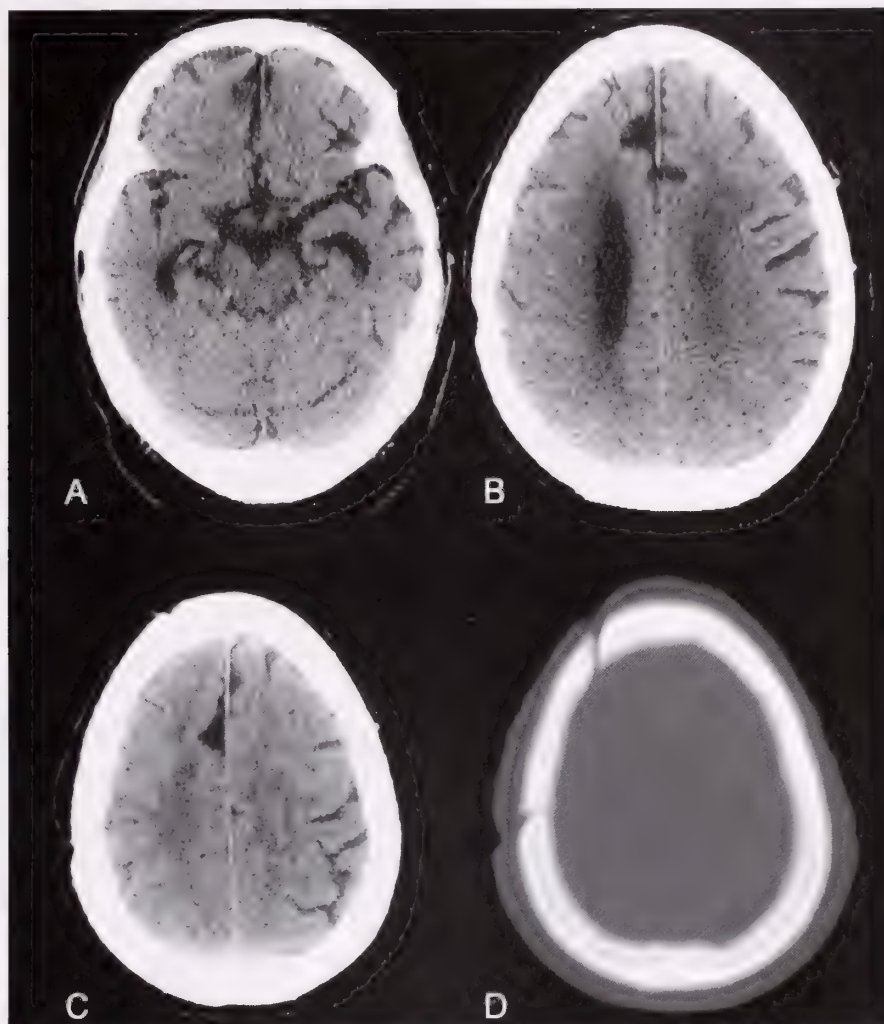


Fig 19. Posttraumatic hydrocephalus. (A,B) The axial computed tomography images show dilatation of the ventricles, including the temporal horns, bilaterally. (C) There is partial effacement of the convexity sulci, which indicates that the enlargement of the ventricles is partially due to hydrocephalus. (D) An image displayed in the bone window shows evidence of a right frontal craniotomy from decompression of a prior intracranial hemorrhage.

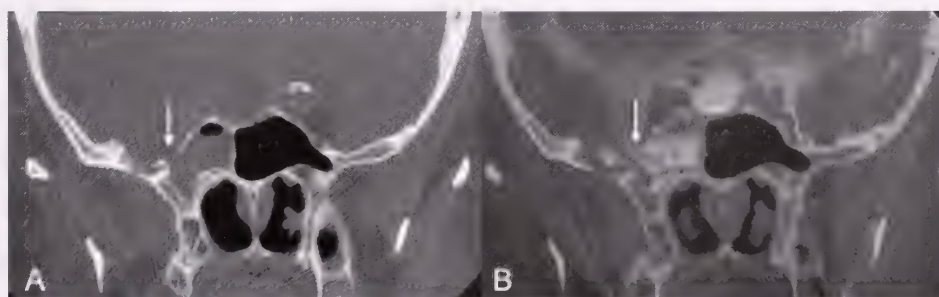


Fig 20. Cerebrospinal fluid leak. (A) The coronal computed tomography image shows a bony defect of the right sphenoid sinus (arrow). (B) The coronal computed tomography image from a cisternogram shows leakage of contrast into the right sphenoid sinus through the bony defect.

16. Noguchi K, Ogawa T, Seto H, et al. Subacute and chronic subarachnoid hemorrhage: diagnosis with fluid attenuated inversion-recovery MR imaging. *Radiology* 1997; 203: 257–262.
17. Woodcock RJ Jr, Short J, Do HM, et al. Imaging of acute subarachnoid hemorrhage with a fluid-attenuated inversion recovery sequence in an animal model: comparison with non-contrast-enhanced CT. *AJNR Am J Neuroradiol* 2001; 22: 1698–1703.
18. Frigon C, Jardine DS, Weinberger E, et al. Fraction of inspired oxygen in relation to cerebrospinal fluid hyperintensity on FLAIR MR imaging of the brain in children and young adults undergoing anesthesia. *AJR Am J Roentgenol* 2002; 179: 791–796.
19. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med* 1999; 42: 952–962.
20. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn Reson Med* 1997; 38: 591–603.
21. Liu AY, Maldjian JA, Bagley LJ, et al. Traumatic brain injury: diffusion-weighted MR imaging findings. *AJNR Am J Neuroradiol* 1999; 20: 1636–1641.
22. Huisman TA, Sorensen AG, Hergan K, et al. Diffusion-weighted imaging for the evaluation of diffuse axonal injury in closed head injury. *J Comput Assist Tomogr* 2003; 27: 5–11.
23. Le TH, Mukherjee P, Henry RG, et al. Diffusion tensor imaging with three-dimensional fiber tractography of traumatic axonal shearing injury: an imaging correlate for the posterior callosal “disconnection” syndrome: case report. *Neurosurgery* 2005; 56: 189.
24. Niogi SN, Mukherjee P, Ghajar J, et al. Extent of microstructural white matter injury in postconcussive syndrome correlates with impaired cognitive reaction time: a 3T diffusion tensor imaging study of mild traumatic brain injury. *AJNR Am J Neuroradiol* 2008; 29: 967–973.
25. Sinson G, Bagley LJ, Cecil KM, et al. Magnetization transfer imaging and proton MR spectroscopy in the evaluation of axonal injury: correlation with clinical outcome after traumatic brain injury. *AJNR Am J Neuroradiol* 2001; 22: 143–151.
26. Lewine JD, Davis JT, Bigler ED, et al. Objective documentation of traumatic brain injury subsequent to mild head trauma: multimodal brain imaging with MEG, SPECT, and MRI. *J Head Trauma Rehabil* 2007; 22: 141–155.
27. Lewine JD, Davis JT, Sloan JH, et al. Neuromagnetic assessment of pathophysiologic brain activity induced by minor head trauma. *AJNR Am J Neuroradiol* 1999; 20: 857–866.
28. Newton MR, Greenwood RJ, Britton KE, et al. A study comparing SPECT with CT and MRI after closed head injury. *J Neurol Neurosurg Psychiatry* 1992; 55: 92–94.
29. Bergsneider M, Hovda DA, McArthur DL, et al. Metabolic recovery following human traumatic brain injury based on FDG-PET: time course and relationship to neurological disability. *J Head Trauma Rehabil* 2001; 16: 135–148.
30. Gross H, Kling A, Henry G, et al. Local cerebral glucose metabolism in patients with long-term behavioral and cognitive deficits following mild traumatic brain injury. *J Neuropsychiatry Clin Neurosci* 1996; 8: 324–334.
31. Alavi A. Functional and anatomic studies of head injury. *J Neuropsychiatry Clin Neurosci* 1989; 1: S45–S50.
32. Al-Nakshabandi NA. The swirl sign. *Radiology* 2001; 218: 133.
33. Greenberg J, Cohen WA, Cooper PR. The “hyperacute” extraaxial intracranial hematoma: computed tomographic findings and clinical significance. *Neurosurgery* 1985; 17: 48–56.
34. Gean AD, Kates RS, Lee S. Neuroimaging in head injury. *New Horiz* 1995; 3: 549–561.
35. Jennett B, Adams JH, Murray LS, Graham DI. Neuropathology in vegetative and severely disabled patients after head injury. *Neurology* 2001; 56: 486–490.
36. Inglese M, Makani S, Johnson G, et al. Diffuse axonal injury in mild traumatic brain injury: a diffusion tensor imaging study. *J Neurosurg* 2005; 103: 298–303.
37. Mittl RL, Grossman RI, Hiehle JF, et al. Prevalence of MR evidence of diffuse axonal injury in patients with mild head injury and normal head CT findings. *AJNR Am J Neuroradiol* 1994; 15: 1583–1589.
38. Scheid R, Preul C, Gruber O, et al. Diffuse axonal injury associated with chronic traumatic brain injury: evidence from T2*-weighted gradient-echo imaging at 3 T. *AJNR Am J Neuroradiol* 2003; 24: 1049–1056.
39. Adams JH, Graham DI, Gennarelli TA, Maxwell WL. Diffuse axonal injury in non-missile head injury. *J Neurol Neurosurg Psychiatry* 1991; 54: 481–483.
40. Ashikaga R, Araki Y, Ishida O. MRI of head injury using FLAIR. *Neuroradiology* 1997; 39: 239–242.
41. Hesselink JR, Dowd CF, Healy ME, et al. MR imaging of brain contusions: a comparative study with CT. *AJR Am J Roentgenol* 1988; 150: 1133–1142.
42. Arfanakis K, Haughton VM, Carew JD, et al. Diffusion tensor MR imaging in diffuse axonal injury. *AJNR Am J Neuroradiol* 2002; 23: 794–802.
43. Gentry LR. Head trauma. In: Atlas SW, ed. *Magnetic Resonance Imaging of the Brain and Spine*. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1996: 611–647.
44. Gean AD. *Imaging of Head Trauma*. Philadelphia, PA: Williams & Wilkins-Lippincott; 1994: 76.
45. Le TH, Gean AD. Imaging of head trauma. *Semin Roentgenol* 2006; 41: 177–189.

Neurobehavioral Sequelae of Traumatic Brain Injury

Silvana Riggio, MD,^{1,2} and Meredith Wong, MD¹

¹Department of Psychiatry, Mount Sinai School of Medicine, New York, NY

²Department of Neurology, Mount Sinai School of Medicine, New York, NY

ABSTRACT

The neurobehavioral sequelae of traumatic brain injury consist of a spectrum of somatic and neuropsychiatric symptoms. The neuropsychiatric symptoms are divided into cognitive and behavioral presentations. In the literature, these neurobehavioral sequelae have been called postconcussive symptoms, postconcussive syndrome, and postconcussive disorder; however, the authors of this review do not use this terminology because the symptoms are not restricted to patients with concussion but instead can be found in all traumatic brain injury patients of all injury severities. The development of neurobehavioral sequelae after traumatic brain injury is a multifactorial process. The patient evaluation requires a multidisciplinary approach in order to delineate physiologic dysfunction and place deficits in the context of the patient's preinjury and postinjury psychiatric status. Consequently, the evaluation of the posttraumatic brain injury patient with neurobehavioral sequelae requires a carefully structured history and physical examination with an emphasis on neurological and psychiatric function. Adjunctive evaluations must be tailored to the patient with neuroimaging, neurophysiological, and neuropsychiatric testing. Maximized outcomes may be achieved by the performance of a careful and detailed assessment that places complaints within the context of the individual. *Mt Sinai J Med* 76:163–172, 2009. © 2009 Mount Sinai School of Medicine

Key Words: frontal lobe seizures, neuropsychiatric disorders, postconcussive syndrome, traumatic brain injury.

It has been estimated that there are 1.4 million new cases of traumatic brain injury (TBI) seen in medical facilities each year in the United States; there are many more who never seek care.¹ The majority of these cases are classified as mild TBI, although regardless of severity, a significant number of TBI patients report some degree of neurobehavioral sequelae (NBS) after the injury.^{2,3} NBS can occur after any type of TBI and are manifest by somatic and/or neuropsychiatric symptoms (see Table 1). Because NBS are reported after all types of TBI, not just after concussion, this article does not use the terms *postconcussive symptoms* and *postconcussive disorder*. In addition, the symptom complex after TBI has been called a postconcussive syndrome; however, this would imply a consistent linkage of symptoms and an underlying pathogenic process that to date has not been demonstrated.

NBS after TBI vary in severity, length, and clinical manifestations. The evolution or resolution of the symptoms over time is multifactorial in nature, with large variability between patients. Symptoms vary, depending on the localization and lateralization of the injury, extent of the injury, medical and psychiatric comorbidities, and prepsychosocial and postpsychosocial factors. Complicating matters even more are the potential side effects of concomitant pharmacotherapy. The relative degree of each of the aforementioned factors in the presentation of NBS is the primary challenge facing the clinician caring for these patients. Therefore, evaluating a patient with NBS must begin with a detailed history that includes premorbid medical, neurological, and psychiatric conditions. The history must also include the use of drugs and medications. Baseline occupational and social function must be established as well as psychological and social stressors that may complicate the post-TBI course. A treatment plan can be developed only once there is a solid understanding of the various factors contributing to the clinical presentation. The goal of this article is to provide an overview of the NBS of TBI and

Address Correspondence to:

Silvana Riggio

Department of Psychiatry

Mount Sinai School of

Medicine, New York, NY

Email: silvana.riggio@mssm.edu

Table 1. *Postconcussive Symptoms.*

Neuropsychiatric
• Cognitive (eg, deficits in attention, memory, and executive function)
• Behavioral
• Primary psychiatric disorder (eg, mood disorder and anxiety)
• Personality disorder
• Other
Somatic (eg, sleep disturbance, fatigue, dizziness, vertigo, headaches, visual disturbances, nausea, sensitivity to light and sound, hearing loss, and seizures)

related factors that enable the development of a comprehensive management strategy.

METHODOLOGY

A MEDLINE literature search was performed of English language publications spanning the period of 1966 to September 2008. Key words included in the search were traumatic brain injury, postconcussive symptoms, postconcussive syndromes, and neurobehavioral sequelae. A total of 259 article abstracts were reviewed, and full-text reviews were performed on selected articles. The references from the full-text articles were scanned for additional relevant studies. The Cochrane Database of Systematic Reviews and the National Guideline Clearinghouse (www.guidelines.gov) were also searched but failed to provide additional relevant material.

Interpretation of the literature on the NBS of TBI is challenging at best because of significant design flaws in the vast majority of studies. Definitions vary from study to study, particularly in those studies examining mild TBI: few studies define the timing of the Glasgow Coma Scale (GCS) score determination or the percentages of patients with GCS scores of 13, 14, or 15. Other design flaws complicating most studies include a failure to document neuroimaging findings. Consequently, a patient with a GCS score of 15 and a negative computed tomography scan could be compared to one with a GCS score of 13 with an acute intracranial lesion; intuitively, the outcomes in these 2 groups would be different, but this is unproven. Additional design flaws commonly encountered when one is trying to analyze this literature include retrospective analyses working off nonstandardized databases, selection bias, small sample sizes, and recall bias. However, perhaps the most glaring shortcomings of the majority of the studies reviewed were (1) the failure to use a control group (ie, a group with no TBI and/or a group with trauma but no TBI) and

(2) no assessment of premorbid mental and physical health. Consequently, almost every article reviewed provided at best a weak level of evidence, and this made definitive conclusions nearly impossible.

PATHOPHYSIOLOGY

Lesions that occur in TBI are classified as the result of either primary or secondary injury. The primary injury is the immediate insult, whereas the secondary injury is due to the biomolecular and physiological changes that follow the insults.⁴ Injury can be focal or diffuse: sudden deceleration or rotational acceleration injury, which generates shearing forces with consequent diffuse axonal injury within the brain, has been implicated as a significant etiology of post-TBI symptoms.⁵ Cortical contusion can result in a loss of function served by a given brain area. White matter lesions can result in interruption of information being transmitted between cortical areas within the brain. Diffuse axonal injuries can result in slowed and inefficient information processing. The impact of injury on neurotransmitter function is poorly defined but clearly could provide a biological explanation for some of the behavioral changes seen after TBI. Table 2 presents behavioral presentations that have been associated with injury to various parts of the brain.

CLINICAL PRESENTATIONS

NBS after TBI are divided into 2 main categories: neuropsychiatric and somatic. The neuropsychiatric

Table 2. *Behavioral Presentation Correlated to Anatomic Brain Injury to Frontal Lobe or Temporal Lobe Area (Most Vulnerable Areas in Posttraumatic Injury).*

Dorsolateral frontal region: Injury may be expressed as difficulties in switching parameters and planning; a certain mental inflexibility can be noted that can ultimately result in irritability, slowness in performance, and/or low frustration tolerance with potential social and performance repercussions.
Orbitofrontal region: Injury can manifest clinically with agitation, disinhibition, and/or poor impulse control.
Medial frontal region: Injury can manifest itself with apathy, which can be misdiagnosed with major depression.
Temporal region: Injury may cause memory disturbance and/or emotional lability problems.
Basal ganglia (or dorsolateral frontal region): Injury may result in mood symptoms (eg, depression), resting tremor, cogwheeling, or bradykinesia.
Right hemispheric limbic area: Injury may result in mania. Lesions to the right and left hemispheres can manifest as psychotic symptoms.

symptoms are subdivided into cognitive and behavioral disorders. Almost all TBI patients report some NBS in the acute phase after injury.⁶ It is estimated that 30% to 80% of patients with mild-to-moderate TBI will experience symptoms of NBS for up to 3 months.^{7,8} Use of different study populations and various definitions contribute to the differences in the reported incidence of NBS. Some of the risk factors that have been identified for the development of NBS include female gender, advanced age, pain, and prior affective or anxiety diagnoses.⁹

In approximately 15% of mild TBI patients, NBS persist beyond 3 months and may contribute to long-term social and occupational difficulties.¹⁰⁻¹² Cognitive dysfunction in the form of impaired attention, memory, and executive function has a predominant role in patients who experience persistent symptoms.⁶ Identifying these deficits and developing an effective intervention plan may be critical to the successful recovery of the patient.

Neuropsychiatric Symptoms

Neuropsychiatric sequelae after TBI can be divided into 2 broad categories: cognitive deficits and behavioral disorders. Significant overlap exists between these sequelae and the somatic manifestations that are presented in the next section. In particular, seizures of frontal lobe origin are categorized as somatic events, and yet their manifestation may be primarily behavioral.

Cognitive Disorders

Cognitive dysfunction is characterized by impairment of attention/concentration, memory, and/or executive function. Patients may have difficulties performing preinjury tasks and jobs or following instructions that would ordinarily be routine. There may be difficulties with sustaining attention, planning, switching parameters, organizing, or sequencing. As a result of these deficits, patients may become irritable, anxious, apathetic, or depressed. Clinical expression may be misinterpreted as secondary to a primary affective disorder with consequent misguided management.

An evaluation of post-TBI cognitive function with a focus on assessing attention versus memory is essential. If attention is impaired, there will be difficulty with retaining information with an obvious impact on memory and thus performance. If the patient has an underlying affective disorder, attention can also be impaired because of a lack of interest and/or distractibility. Therefore, the assessment of memory must be placed in the context of attention, and a detailed psychiatric history is warranted to

exclude other disorders that may interfere with performance.

Cognitive deficits in mild TBI generally resolve within days and rarely last more than 3 months.^{13,14} Loss of consciousness (LOC) and posttraumatic amnesia (PTA) have historically been considered markers for injury; however, a growing body of literature supports cognitive deficits even when LOC or PTA does not occur. A prospective study of cognitive functioning used pre-TBI assessments of 91 high school and college football players and compared their performance before and after mild TBI during the season. The authors reported cognitive impairment with respect to the athletes' own and matched control baselines immediately after TBI, even in the absence of LOC or PTA.¹⁴ Greater impairment was seen in those with LOC, however, and this suggested a gradient of concussion severity with respect to LOC and PTA. Significant impairment was detected at 15 minutes, but they returned to baseline levels of cognitive functioning within 48 hours.

Behavioral Disorders

Behavioral manifestations after TBI can be due to a psychiatric disorder or may be secondary to other processes, such as cognitive problems or medical or structural lesions. The challenge for the clinician is to determine if the symptoms were present before the injury and the role of the injury in the symptom, that is, the primary cause or an exacerbant. Table 2 presents a summary of behavioral manifestations related to injury to various parts of the brain. In addition, symptoms may be secondary to an emotional response to the injury, its physical limitations, or fear of the impact on function. Weak defense mechanisms, poor social support, medication, and drug use can all complicate the presentation.

Personality Changes Affective and behavioral disturbances after TBI may be expressed as personality changes appreciated by the patients or their families/caregivers. Personality changes may include aggression, impulsivity, irritability, emotional lability, and apathy.^{15,16} Impulsivity and irritability may lead to verbal and physical inappropriateness expressed as verbal outbursts or combativeness. This may be due to impaired judgment secondary to an underlying structural lesion or the exacerbation of an underlying psychiatric disorder or to an emotional response to trauma.

Aggression is a commonly reported behavioral symptom of TBI. Tateno *et al.*¹⁵ studied 89 consecutive inpatients with moderate-to-severe TBI and

found that 34% exhibited aggressive behavior within the first 6 months of injury. Risk factors for aggression after TBI include frontal lobe injury, premorbid affective disorder, personality disorder, and alcohol or substance abuse.

Major Depression Major depression is one of the most frequently reported behavioral sequelae of TBI; the actual prevalence varies from study to study according to the methodology but appears to be approximately 25% to 40% after moderate-to-severe TBI versus a general population prevalence of 17%.^{17,18} The degree to which a premorbid psychiatric disorder increases the risk for NBS after TBI is unclear, but studies indicate a positive correlation. Seel *et al.*¹⁹ reported fatigue (29%), distractibility (28%), anger/irritability (28%), and rumination (25%) to be the most common depressive symptoms in a prospective, multicenter study of 666 nonacute moderate-to-severe TBI patients. Twenty-seven percent of the TBI patients met criteria for major depression, with feelings of hopelessness, worthlessness, and anhedonia differentiating depressed patients from nondepressed patients.

Risk factors for developing major depression after TBI fall into 2 categories: premorbid psychiatric pathology and low socioeconomic status. The relationship between rates of depression and the severity of TBI is unclear. However, the greatest chronicity of major depression appears to occur in patients with both moderate-to-severe TBI and prior psychiatric illness.

Studies have found a link between TBI and suicidality and between psychiatric comorbidity in the setting of TBI and suicidality.^{20,21} In a retrospective study of 5034 patients, Silver *et al.*²⁰ reported that patients with a history of TBI with LOC posed a 4 times greater likelihood of attempted suicide than those without TBI: 8.1% versus 1.9%.²⁰ This risk of suicide attempt remained even after they controlled for demographics, quality of life variables, alcohol abuse, and any comorbid psychiatric disorders.

Anxiety Disorders/Posttraumatic Stress Disorder An estimated 24% of mild TBI patients develop a new anxiety disorder after their injury versus a population-based prevalence of approximately 7%.²² Increased age, a history of posttraumatic stress disorder (PTSD), and an avoidant coping style increase the risk of acute stress symptoms after TBI.^{23,24} In turn, a diagnosis of acute stress disorder is a risk factor for the development of PTSD after TBI. In a study of 79 patients with mild TBI,

Bryant and Harvey²⁵ diagnosed 14% of the patients with acute stress disorder at 1 month, and 24% were diagnosed with PTSD at 6 months post-injury; 82% of the patients diagnosed with acute stress disorder had developed PTSD by 6 months.

Another predictor of PTSD after TBI is first-hand recall of the traumatic event. In a study by Gil *et al.*²⁶ 14% of 120 patients with mild TBI developed PTSD by 6 months: 45% of these patients recalled the event and 55% did not recall the event, and recall of the event was shown to be a strong predictor of PTSD diagnosis. Turnbull *et al.*²⁷ in a study of 55 patients showed that having traumatic memories or no memories of the event was associated with increased psychological distress. On the other hand, nontraumatic memories were associated with good psychological adjustment. Having amnesia for the event, although not protective against PTSD, was related to a decreased severity of symptoms and was specifically protective against intrusive PTSD symptoms. Other predictors of the development of PTSD as well as its severity include an avoidant coping style, a behavioral coping style (versus a cognitive coping style), and prior unemployment (an indicator of the premorbid level of functioning).²⁸

There exists controversy over whether or not PTSD is truly diagnosable in the case of PTA, in which the patient does not recall the traumatic event. Opponents of the ability to diagnose state that the patient cannot technically have PTSD because they are unable to re-experience the actual trauma via intrusive thoughts or nightmares.^{29,30} Proponents of the ability to diagnose posit that TBI patients with PTA for the traumatic event can re-experience vivid pseudomemories of the event generated by the imagination from second-hand accounts or the patient's own mind³¹ and that patients may also be able to re-experience fragments or islands of memory within the amnesic period,³² thus meeting the re-experiencing criterion of PTSD.

Substance Use Disorders A review of the literature by van Reekum *et al.*¹⁸ reported a 22% prevalence of substance abuse in TBI patients versus a 15% lifetime prevalence in the general population. A review of subsequent studies by Rogers and Read¹⁷ in 2007 showed a prevalence of 12%. Premorbid substance use has been found to be strongly associated with post-TBI drug use, and multiple studies have cited substance abuse as a risk factor for TBI rather than the other way around. A 30-year longitudinal study by Koponen *et al.*³³ showed that 71% of TBI patients who were currently using drugs also did so pre-TBI.

Somatic Symptoms

Headache

Headache is the most commonly reported somatic symptom after mild traumatic brain injury.³⁴ The prevalence of post-TBI headache varies greatly by study, ranging from 25% to 90% of patients.^{35,36} Post-TBI headaches may be classified as acute or chronic. According to the International Headache Society, acute posttraumatic headaches begin within 2 weeks of the injury and resolve within 2 months; chronic posttraumatic headaches begin within 2 weeks and persist for more than 8 weeks.³⁵ The common co-occurrence of headache with other NBS was noted by Baandrup and Jensen³⁵: 53% had at least 1 other somatic complaint (fatigability, sleep disturbance, dizziness, or alcohol intolerance), 49% had at least 1 cognitive complaint (memory dysfunction or impaired concentration/attention), 26% had at least 1 psychiatric complaint (irritability, aggressiveness, anxiety, depression, or emotional lability), 17% had all 3 types of complaints, and 17% had none.

A premorbid history of headache increases the risk of posttraumatic headaches, and those with pretraumatic headache may experience a worsening of the headache after injury. The presence of posttraumatic headache has not been consistently correlated with the severity of the injury; in fact, some authors have reported that mild TBI patients have higher rates of headache during the initial posttraumatic phase than patients with more severe injury.³⁷ In the majority of mild TBI cases, headaches resolve within 3 to 6 months.⁴⁸

Dizziness/Nausea

Dizziness is the second most commonly reported somatic symptom of postconcussion syndrome.³⁹ Most studies do not differentiate posttraumatic dizziness from vertigo, although the pathophysiology may be greatly different. Vertigo, characterized by the appearance of movement of the environment around oneself, may be peripheral or central in etiology. Peripheral etiologies include cupulolithiasis, perilymphatic fistula, posttraumatic Meniere's disease, damage to the vestibular nerve, and use of ototoxic medications. Central etiologies include damage to the brainstem involving the vestibular nucleus. Dizziness/vertigo is reported in 24% to 78% of mild TBI patients acutely, and this is significantly higher than the prevalence in non-TBI patients in the community.⁴⁰ As with headache, dizziness generally resolves within 6 months in patients with mild TBI, whereas persistence may be more problematic in patients with moderate or severe TBI.⁴¹

Fatigue

Fatigue is reported in up to 73% of patients post-TBI and may become a persistent and debilitating symptom.⁴² The presence of fatigue is associated with poorer social integration, decreased levels of productive activities, and decreased overall quality of life.⁴³ When fatigue persists, it may present a barrier to recovery. Severity of TBI and age have not been found to be predictors of severity of fatigue.⁴⁴ Post-TBI fatigue is most likely the result of a combination of etiologies. Studies have shown that fatigue can be associated with pain,⁴³ sleep disorders,⁴⁵ cognitive deficits,⁴⁶ and depression and anxiety.⁴³ Hypopituitarism, with resultant neuroendocrine abnormalities such as growth hormone deficiency and cortisol deficiency, may also be associated with post-TBI fatigue.⁴⁷ Other possible contributing factors include vertigo, diplopia, and iatrogenic causes, such as psychotropic or analgesic medications.

Sleep Disturbance

Sleep disturbances include difficulties in initiating sleep, maintaining sleep, or attaining restful sleep as well as excessive daytime somnolence and, less commonly, parasomnias. It is reported in up to 73% of post-TBI patients, and this is greater than the 32% to 35% prevalence reported in the general population.⁴⁸ Sleep disturbance has not been clearly linked to severity of TBI.⁴⁹ Abnormalities on polysomnography in mild TBI patients with chronic sleep disturbance have been shown, and as with all the other somatic symptoms, the etiology is complex and therefore takes more than a prescription to solve.⁵⁰

Seizures

Posttraumatic seizures can be focal or generalized; focal seizures can be simple or complex according to the lack (simple) or presence (complex) of a loss of awareness or consciousness. Both simple and complex partial seizures can evolve into a secondary generalized seizure with tonic clonic activity and postictal confusion. Because of the high incidence of TBI to the frontal or temporal lobe, the focus of posttraumatic seizures is often in these areas. Seizures of frontal lobe origin can be difficult to diagnose as they can present with a wide variety of clinical manifestations, frequently mimicking psychiatric disorders, yet can have a normal electroencephalogram (EEG).⁵¹

Seizures of frontal origin are often characterized by bizarre clinical manifestations such as repetitive motor activity, simple or complex automatisms

Table 3. *Clinical Characteristics Differentiating Frontal Lobe Seizures from Other Causes of Bizarre/Abnormal Behavior.*

Presentation	Frontal Lobe Seizure	Nonseizure Disorder
Stereotypicity	Yes	No
RMA	Yes	No
Aura	Possible	No
Duration	Brief	Variable but usually prolonged
Acute onset	Yes	Variable
Abrupt ending	Yes	No
Postictal phase	Possible	No
Frequency	Often in clusters	Variable
Nocturnal	Possible	No
Situational	No	Yes
Goal-directed aggression	No	Yes

Abbreviation: RMA, repetitive motor activity.

(trashing, kicking, tapping, running, and agitation), and/or nonspecific events including dizziness, light-headedness, and delusions.⁵¹ The bizarreness of frontal lobe seizures, without associated tonic-clonic activity or LOC, results in the diagnosis frequently being overlooked. To compound this issue, much of the frontal lobe is inaccessible to standard scalp EEG recording; consequently, the EEG can be normal even during an ictal episode. This fact can result in the false exclusion of a seizure disorder when too much reliance is placed on a single diagnostic test. It is important to remember that standard EEG in any type of focal epilepsy detects interictal epileptiform abnormalities in only 29% to 55% of patients and that several series have reported no ictal EEG change in 33% to 36% of patients.⁵² Some series have reported that only 14% of patients with frontal lobe seizures have localized frontal lobe discharges.⁵¹

Table 3 presents clinical characteristics that help differentiate frontal lobe seizures from other disorders causing abnormal behavior. When agitation, brief outbursts of rage and aggression, new onset bizarre behavior, or a combination of these occurs in association with a TBI, a detailed history and careful characterization of the events are warranted. If the episodes are periodic, stereotypic, and brief in duration with a variable postictal phase, then the possibility of a frontal lobe seizure should be entertained.

DISCUSSION

The etiology of NBS of TBI is a complex interface between structural injury, psychiatric predisposition,

Table 4. *Key Points.*

Neurobehavioral sequelae after TBI are divided into somatic and neuropsychiatric symptoms. Neurobehavioral sequelae after TBI are multifactorial in etiology and require a careful, systematic evaluation of physical, neurological, and psychiatric function. Therapy for neurobehavioral sequelae must be tailored to the individual patient and based on a coordinated, comprehensive evaluation. Headache is the most common somatic complaint after TBI, and depression is the most common psychiatric complaint. Seizures of frontal lobe origin can manifest as bizarre behavior; diagnosis requires a high index of suspicion with an understanding that the EEG may be normal.

Abbreviations: EEG, electroencephalogram; TBI, traumatic brain injury.

and environmental-social-economic factors. These processes are not exclusive, so a coordinated multidisciplinary approach is needed to develop an effective diagnostic and management strategy. Symptoms that can be debilitating for one person might not have the same impact on someone else. Figure 1 schematically presents the various factors that must be considered in assessing post-TBI patients with NBS.

Intuitively, it would seem that the greater the severity of injury, the more structural damage to the brain and the greater the NBS, but this is not as clearly demonstrated as one would assume. On the mild end of the spectrum (GCS score: 13–15), acute intracranial lesions, as demonstrated on computed tomography, occur less frequently than in patients with a GCS score less than 13, and this makes the diagnosis of injury more challenging. A number of studies have been performed to delineate the etiology and establish the existence of NBS in mild TBI; the findings have been controversial and inconclusive. Hoge *et al.*⁵³ studied soldiers with mild TBI and a variety of postinjury symptoms. After adjustments for PTSD and depression, mild TBI was no longer significantly associated with symptoms except headache. The authors postulated that PTSD and depression are important mediators of the relationship between mild TBI and physical health problems. In another provocative study, Iverson and McCracken⁵⁴ studied the prevalence of post-TBI symptoms in patients with non-TBI chronic medical conditions: they reported that 94% of these patients met criteria for commonly ascribed TBI-related symptoms. They reported disturbed sleep, fatigue, and/or irritability in 81% of the patients and 1 or more cognitive problems in 42% of the patients.

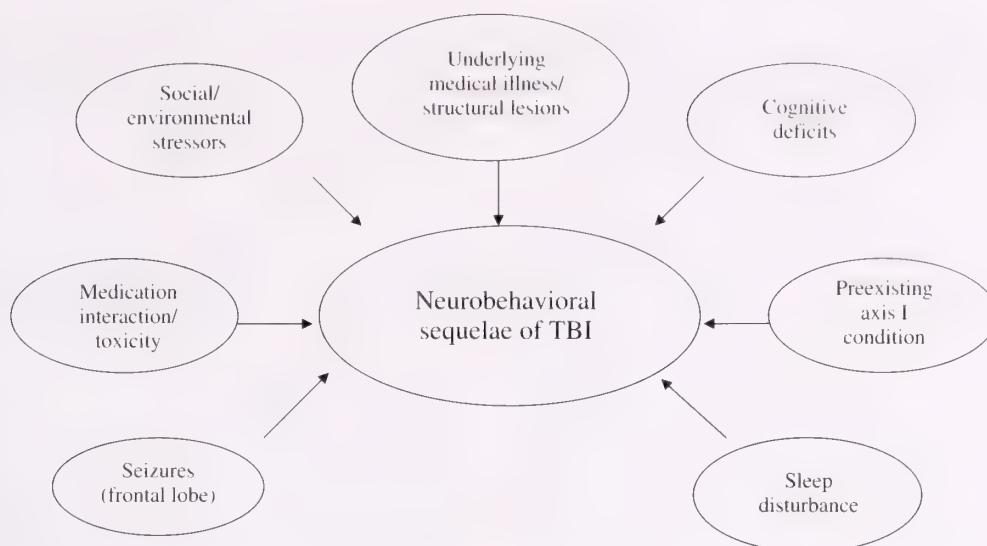


Fig 1. Assessment scheme for evaluating changes in behavior after traumatic brain injury (TBI).

Other authors have reported similar findings.^{9,23,55,56} Meares *et al.*⁹ studied a sample of consecutive trauma patients admitted to a level 1 trauma center, including 90 patients with mild TBI and 85 with non-brain injury trauma: both groups had the same incidence of symptoms, with the strongest predictor of symptoms in either group being a previous affective disorder. A correlation between pain and NBS has been reported, and pain has been associated with the persistence of symptoms.^{9,56} Hart *et al.*⁵⁶ reported that pain after TBI was associated with cognitive impairment, including deficits in attention, memory, processing speed, and reaction time. The occurrence of cognitive complaints in non-TBI chronic pain patients has been demonstrated, and this once again questions the relationship between TBI per se and NBS.⁵⁷

The previously cited studies demonstrate the complexity of assessing post-TBI, especially after mild TBI. However, it must be emphasized that they do not negate the potential for TBI to directly result in significant disability. At the risk of stating the obvious, we note that the evaluation of a post-TBI complaint must begin with a systematic, comprehensive physical, neurological, and psychiatric examination. For example, a complaint of persisting headache cannot be diagnosed generically as post-TBI headache until structural lesions, including subtle cranial nerve dysfunction, are excluded. Complaints of depression, inattention, and bizarre behavior provide additional examples of the complexity of assessing post-TBI NBS.

Depressed mood might not necessarily be a major depressive disorder but instead might be the result of an adjustment disorder, a lesion in the mesial frontal region, an injury to the basal ganglia,

or an undiagnosed hypothyroid state: the distinction is critical. An adjustment disorder might not require pharmacotherapy, whereas a major depression disorder does; a patient with depression from a mesial frontal lesion would obtain questionable benefit from antidepressants. Depression from a basal ganglia lesion could potentially get worse with antidepressants because of the possibility of secondary extrapyramidal symptoms, with bradykinesia mimicking psychomotor retardation. Hypothyroidism requires hormone replacement.

Many of the NBS can be linked to the impact of TBI on attention, and this makes assessing attention a key component of the NBS evaluation. Attention deficits may be due to structural lesions, an underlying medical problem, a primary psychiatric disorder, or drug side effect (in particular, compounds with high anticholinergic properties or antihistaminergic properties). Sleep disturbances and pain syndromes can also interfere with attention. Attention can be affected by a lesion in the orbitofrontal region resulting in high distractibility or by a lesion in the dorsolateral region resulting in an inability to switch mind set or to multitask. Mood disorder, anxiety disorder, psychotic disorder, and/or personality disorder can also interfere with attention, as can fatigue secondary to the injury or to a medical problem.

Bizarre or unusual behavior after TBI can be the result of structural lesions, psychiatric disorders, drug toxicity, metabolic disorders, or frontal lobe seizures. Frontal lobe seizures deserve special mention because, as discussed in the section on seizures, their clinical presentation can be characterized by bizarre behavior, such as repetitive motor

activity, or automatisms, such as repetitive tapping, kicking, running, or trashing. These episodes can be associated either with or without a loss of awareness, no LOC, and variable postictal confusion. Findings can appear psychiatric in origin, and presentation can be associated with a normal interictal or ictal EEG (see Table 3).

The key to the diagnosis and management of post-TBI complaints is to avoid premature closure on a diagnosis, to coordinate care through a multidisciplinary team, and to involve the patient and his or her family in decision making. There is evidence, primarily from the mild TBI literature, supporting the benefit of education and reassurance after TBI on outcome. Ponsford *et al.*¹¹ studied 202 mild TBI patients and reported that patients given an information booklet on mild TBI and coping strategies for symptoms were significantly less symptomatic at 3 months than those who were not provided education. An extensive review of articles on early intervention after mild TBI by Borg *et al.*⁵⁸ showed that early educational information reduces long-term complaints.

CONCLUSIONS

The NBS of TBI are composed of a spectrum of somatic and neuropsychiatric symptoms. Controversy exists regarding the structural/biological etiology versus the psychiatric etiology of NBS, most likely because the cause is rarely within 1 domain. Clinical presentations must be placed in the context of the patient's premorbid state. Injury localization can be difficult, and thus a critical component of the patient's evaluation is a meticulously performed physical, neurological, and psychiatric examination. Because of the limitations of neuroimaging, additional assessment may require neurophysiological and neuropsychological testing. Effective management strategies must be tailored to the individual patient, based on an understanding of the symptom etiology, and placed within the patient's social and functional framework. Key point related to the diagnosis and management of neurobiological sequelae of TBI are presented in Table 4.

There is a need for continued research in order to better understand NBS seen after TBI. Regardless of the etiology, the symptoms observed contribute to significant disability for many patients. The key is to maximize recovery and minimize confounding the presentation by the use of unnecessary medications or other interventions. Future research will benefit from well-designed studies that are adequately

powered and that include appropriate control groups. The incidence of TBI is increasing worldwide, and this is creating an urgency to delineate best practice strategies needed to improve outcomes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Division of Injury and Disability Outcomes and Programs, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Department of Health and Human Services. Traumatic brain injury in the United States: emergency department visits, hospitalizations, and deaths, October 2004. Available at: <http://www.cdc.gov/ncipc/pub-res/tbi.in.us.04/tbi.ed.htm>. Accessed January 2009.
2. Haas DC. Chronic post-traumatic headaches classified and compared with natural headaches. *Cephalalgia* 1996; 16: 486–490.
3. McAllister T, Arciniegas D. Evaluation and treatment of postconcussive symptoms. *NeuroRehabilitation* 2002; 17: 265–283.
4. Raghupathi R. Cell death mechanisms following traumatic brain injury. *Brain Pathol* 2004; 14: 215–222.
5. Povlishock JT. Pathobiology of traumatically induced axonal injury in animals and man. *Ann Emerg Med* 1993; 22: 980–988.
6. Lundin A, de Boussard C, Edman G, et al. Symptoms and disability until 3 months after mild TBI. *Brain Inj* 2006; 20: 799–806.
7. Rimel RW, Giordani B, Barth JT, et al. Disability caused by minor head injury. *Neurosurgery* 1981; 9: 221–228.
8. McCauley SR, Boake C, Pedroza C, et al. Postconcussional disorder: are the DSM-IV criteria an improvement over the ICD-10? *J Nerv Ment Dis* 2005; 193: 540–550.
9. Meares S, Shores EA, Taylor AJ, et al. Mild traumatic brain injury does not predict acute postconcussion syndrome. *J Neurol Neurosurg Psychiatry* 2008; 79: 300–306.
10. Alves W, Macciocchi S, Barth J. Postconcussive symptoms after uncomplicated mild head injury. *J Head Trauma Rehabil* 1993; 8: 48–59.
11. Ponsford J, Willmott C, Rothwell A. Impact of early intervention on outcome following mild head injury in adults. *J Neurol Neurosurg Psychiatry* 2002; 73: 330–332.
12. Ruff RM, Camenzuli L, Mueller J. Miserable minority: emotional risk factors that influence the outcome of a mild traumatic brain injury. *Brain Inj* 1996; 8: 61–65.
13. Bleiberg J, Cernich AN, Cameron K, et al. Duration of cognitive impairment after sports concussion. *Neurosurgery* 2004; 54: 1073–1080.
14. McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA concussion study. *JAMA* 2003; 290: 2556–2563.

15. Tateno A, Jorge RE, Robinson RG. Clinical correlates of aggressive behavior after traumatic brain injury. *J Neuropsychiatry Clin Neurosci* 2003; 15: 155–160.
16. Greve KW, Sherwin E, Stanford MS, et al. Personality and neurocognitive correlates of impulsive aggression in long-term survivors of severe traumatic brain injury. *Brain Inj* 2001; 15: 255–262.
17. Rogers JM, Read CA. Psychiatric comorbidity following traumatic brain injury. *Brain Inj* 2007; 21: 1321–1333.
18. van Reekum R, Cohen T, Wong J. Can traumatic brain injury cause psychiatric disorders? *J Neuropsychiatry Clin Neurosci* 2000; 12: 316–327.
19. Seel RT, Kreutzer JS, Rosenthal M, et al. Depression after traumatic brain injury: a national institute on disability and rehabilitation research model systems multicenter investigation. *Arch Phys Med Rehabil* 2003; 84: 177–184.
20. Silver JM, Kramer R, Greenwald S, et al. The association between head injuries and psychiatric disorders: findings from the New Haven NIMH Epidemiologic Catchment Area Study. *Brain Inj* 2001; 15: 935–945.
21. Simpson G, Tate R. Suicidality in people surviving a traumatic brain injury: prevalence, risk factors and implications for clinical management. *Brain Inj* 2007; 21: 1335–1351.
22. Mooney G, Speed J. The association between mild traumatic brain injury and psychiatric conditions. *Brain Inj* 2001; 15: 865–877.
23. Bryant RA. Posttraumatic stress disorder and traumatic brain injury: can they co-exist? *Clin Psychol Rev* 2001; 21: 931–945.
24. Harvey AG, Bryant RA. Predictors of acute stress following mild traumatic brain injury. *Brain Inj* 1998; 12: 147–154.
25. Bryant RA, Harvey AG. Relationship between acute stress disorder and posttraumatic stress disorder following mild traumatic brain injury. *Am J Psychiatry* 1998; 155: 635–629.
26. Gil S, Caspi Y, Ben-Ari IZ, et al. Does memory of a traumatic event increase the risk of posttraumatic stress disorder in patients with traumatic brain injury? A prospective study. *Am J Psychiatry* 2005; 162: 963–969.
27. Turnbull SJ, Campbell EA, Swann IJ. Post-traumatic stress disorder symptoms following a head injury: does amnesia for the event influence the development of symptoms? *Brain Inj* 2001; 15: 775–785.
28. Bryant RA, Marosszeky JE, Crooks J, et al. Coping style and post-traumatic stress disorder following severe traumatic brain injury. *Brain Inj* 1999; 14: 175–180.
29. Sbordone RL, Liter JC. Mild traumatic brain injury does not produce post-traumatic stress disorder. *Brain Inj* 1995; 9: 405–412.
30. Warden DL, Labbate LA, Salazar AM, et al. Posttraumatic stress disorder in patients with traumatic brain injury and amnesia for the event? *J Neuropsychiatry Clin Neurosci* 1997; 9: 18–22.
31. Bryant RA. Posttraumatic stress disorder, flashbacks and pseudomemories in closed head injury. *J Trauma Stress* 1996; 9: 621–629.
32. King NS. Post-traumatic stress disorder and head injury as a dual diagnosis: “islands” of memory as a mechanism. *J Neurol Neurosurg Psychiatry* 1997; 62: 82–84.
33. Koponen S, Taiminen T, Portin R. Axis I and II psychiatric disorders after traumatic brain injury: a 30-year follow-up study. *Am J Psychiatry* 2002; 159: 1315–1321.
34. Uomoto JM, Esselman PC. Traumatic brain injury and chronic pain: differential types and rates by head injury severity. *Arch Phys Med Rehabil* 1993; 74: 61–64.
35. Baandrup L, Jensen R. Chronic post-traumatic headache—a clinical analysis in relation to the international headache classification 2nd edition. *Cephalalgia* 2005; 25: 132–138.
36. Paniak C, Reynolds S, Phillips K, et al. Patient complaints within 1 month of mild traumatic brain injury: a controlled study. *Arch Clin Neuropsychol* 2002; 17: 319.
37. Couch JR, Bearss C. Chronic daily headache in the posttrauma syndrome: relation to the extent of head injury. *Headache* 2001; 41: 559.
38. Packard RC. Epidemiology and pathogenesis of posttraumatic headache. *J Head Trauma Rehabil* 1999; 14: 9–21.
39. Chamelian L, Feinstein A. Outcome after mild to moderate traumatic brain injury: the role of dizziness. *Arch Phys Med Rehabil* 2004; 85: 1662–1666.
40. Anstey KJ, Butterworth P, Jorm AF, et al. A population survey found an association between self-reports of traumatic brain injury and increased psychiatric symptoms. *J Clin Epidemiol* 2004; 57: 1202–1209.
41. Masson F, Maurette P, Salmi LR, et al. Prevalence of impairments 5 years after a head injury, and their relationship with disabilities and outcome. *Brain Inj* 1996; 10: 487–497.
42. Bushnik T, Englander J, Wright J. The experience of fatigue in the first 2 years after moderate-to-severe traumatic brain injury: a preliminary report. *J Head Trauma Rehabil* 2008; 23: 17–24.
43. Cantor JB, Ashman T, Gordon W, et al. Fatigue after traumatic brain injury and its impact on participation and quality of life. *J Head Trauma Rehabil* 2008; 23: 41–51.
44. Ziino C, Ponsford J. Selective attention deficits and subjective fatigue following traumatic brain injury. *Neuropsychology* 2006; 20: 383–390.
45. Ouellet M, Beaulieu-Bonneau S, Morin CM. Insomnia in patients with traumatic brain injury: frequency, characteristics, and risk factors. *J Head Trauma Rehabil* 2006; 21: 199–212.
46. Ashman TA, Cantor JB, Gordon WA, et al. Objective measurement of fatigue following traumatic brain injury. *J Head Trauma Rehabil* 2008; 23: 33–40.
47. Popovic V. GH deficiency as the most common pituitary defect after TBI: clinical implications. *Pituitary* 2005; 8: 239–243.
48. Rao V, Rollings P. Sleep disturbances following traumatic brain injury. *Curr Treat Options Neurol* 2002; 4: 77–87.
49. Clinchot DM, Bogner J, Mysiw WJ, et al. Defining sleep disturbance after brain injury. *Am J Phys Med Rehabil* 1998; 77: 291–295.
50. Williams BR, Lazic SE, Ogilvie RD. Polysomnographic and quantitative EEG analysis of subjects with long-term insomnia complaints associated with mild traumatic brain injury. *Clin Neurophysiol* 2008; 119: 429–438.
51. Riggio S, Harner RN. Repetitive motor activity in frontal lobe epilepsy. In: Jasper H, Silvana MDR, Goldman-Rakic PS, eds. *Epilepsy and the Frontal Anatomy of*

- the Frontal Lobe*. New York, NY: Raven Press; 1995; 153–166.
52. Quesney LF. Seizures of frontal lobe origin. In: Pedley T, Meldrum B, eds. *Recent Advances in Epilepsy*. Vol 3. London, England: Churchill Livingstone; 1986; 81–100.
 53. Hoge CW, McGurk D, Thomas JL, et al. Mild traumatic brain injury in soldiers returning from Iraq. *N Engl J Med* 2008; 358: 453–463.
 54. Iverson GL, McCracken LM. 'Postconcussive' symptoms in persons with chronic pain. *Brain Inj* 1997; 11: 783–790.
 55. Lees-Haley PR, Brown RS. Neuropsychological complaints base rates of 170 personal injury claimants. *Arch Clin Neuropsychol* 1993; 8: 203–209.
 56. Hart RP, Martelli MF, Zasler ND. Chronic pain and neuropsychological functioning. *Neuropsychol Rev* 2000; 10: 131–149.
 57. McCracken LM, Iverson GL. Predicting complaints of impaired cognitive functioning in patients with chronic pain. *J Pain Symptom Manage* 2001; 21: 392–396.
 58. Borg J, Holm L, Peloso PM, et al. Non-surgical intervention and cost for mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J Rehabil Med* 2004; 43(suppl): 76–83.

Cognitive Rehabilitation Following Traumatic Brain Injury: Assessment to Treatment

Theodore Tsaousides, PhD, and Wayne A. Gordon, PhD

Department of Rehabilitation Medicine, Mount Sinai School of Medicine, New York, NY

ABSTRACT

Cognitive rehabilitation refers to a set of interventions that aim to improve a person's ability to perform cognitive tasks by retraining previously learned skills and teaching compensatory strategies. Cognitive rehabilitation begins with a thorough neuropsychological assessment to identify cognitive strengths and weaknesses and the degree of change in cognitive ability following a brain injury. The conclusions of the assessment are used to formulate appropriate treatment plans. Common interventions for improvements in attention, memory, and executive function, as well as the nature of comprehensive programs, which combine treatment modalities, are reviewed. Cognitive rehabilitation is effective for mild-to-severe injuries and beneficial at any time post-injury. Sufficient evidence exists supporting the efficacy and effectiveness of cognitive rehabilitation, which has become the treatment of choice for cognitive impairments and leads to improvements in cognitive and psychosocial functioning. *Mt Sinai J Med* 76:173–181, 2009. © 2009 Mount Sinai School of Medicine

Key Words: attention,, cognitive deficits,, cognitive rehabilitation,, comprehensive day treatment,, executive function,, memory,, traumatic brain injury.

Traumatic brain injury (TBI) often results in physical, cognitive, and emotional impairments that interfere

with independent living and disrupt psychosocial and vocational functioning.¹ The consequences of TBI are debilitating and persistent, especially when they remain untreated. The aim of rehabilitation following TBI is to improve physical, cognitive, and psychosocial functioning, to foster independence, and to facilitate community integration.² Cognitive rehabilitation is often the treatment of choice for these diverse impairments. Although cognitive rehabilitation targets cognitive and psychosocial functioning more directly, improvements in cognitive functioning could lead to improvements in physical functioning indirectly. For example, improvements in memory may facilitate medication compliance, improvements in attention and comprehension may improve understanding of instructions given by medical personnel, and improvements in executive function may facilitate initiating medical appointments or improve decision-making with respect to treatment options.

The American Congress of Rehabilitation Medicine has defined cognitive rehabilitation as “a systematic, functionally oriented service of therapeutic activities that is based on assessment and understanding of the patient's brain-behavioral deficits” (Cicerone *et al.*,³ p 1597). Thus, as in all medical or psychological interventions, diagnosis and treatment are tied together. The goal of cognitive rehabilitation is to improve the person's ability to perform cognitive tasks, cope with affective distress, and increase self-confidence, self-efficacy, and self-awareness. This is achieved by retraining previously learned skills and residual abilities, teaching the person compensatory strategies, making environmental modifications to the person's domestic and vocational setting, and facilitating adjustment to the cognitive disability by increasing awareness and acceptance. These approaches are often combined to optimize the effects of treatment.³

Address Correspondence to:

Theodore Tsaousides
Department of
Rehabilitation Medicine
Mount Sinai School of Medicine
New York, NY
Email: theodore.tsaousides@
mssm.edu

METHODOLOGY

The purpose of this article is 2-fold: to examine the available literature that supports cognitive rehabilitation as an efficacious and effective treatment following TBI and to provide an overview of commonly used cognitive rehabilitation interventions. The review of the evidence for cognitive rehabilitation is based on 3 systematic reviews that were conducted by a subcommittee of the Brain Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine and 1 review that was commissioned by the National Institute on Disability and Rehabilitation Research. The goal of these reviews was to examine the state of the evidence for cognitive rehabilitation interventions following TBI and stroke, and they have resulted in 3 seminal publications to date.³⁻⁵ The methodology used in these publications, including the search terms used to identify publications, the number of articles identified and retained, and the level of evidence classification, is described in a subsequent section.

To accomplish the second goal, to provide an overview of common cognitive rehabilitation interventions, among the studies included in the published reviews,³⁻⁵ those that focused on TBI were located and reviewed. The description of the cognitive rehabilitation interventions that follows is thus based on class I and class II studies with TBI samples.

ROLE OF ASSESSMENT PRIOR TO COGNITIVE REHABILITATION

Neuropsychological assessment is the initial part of cognitive rehabilitation and serves several purposes. It identifies areas of cognitive function in need of treatment as the data that result from the assessment are used to make inferences about the nature and extent of the person's cognitive dysfunction. Awareness of the person's cognitive strengths and weakness provides a means of targeting the cognitive domains that will require remediation and capitalizing on the person's residual cognitive abilities to facilitate treatment. In addition, neuropsychological assessment provides a means for evaluating the effectiveness of treatment.⁶

Neuropsychological assessment examines several domains of cognitive function. The individual measures included in a neuropsychological battery are determined on a case-by-case basis according to the patient's self-report of his day-to-day cognitive failures and clinical observation. The most

common domains of cognitive functioning that are assessed post-TBI include intellectual function, memory function, psychomotor speed, processing speed, attention, language, and executive function.

General Intellectual Functioning

Neuropsychological assessment involves an evaluation of the individual's general intellectual functioning. Tests of intellectual function examine the performance of diverse mental functions, including attention, processing speed, visual-spatial perception and construction, concept formation and abstraction, judgment, verbal comprehension, and fund of knowledge. Assessing general intellectual functioning post-TBI provides the clinician with a means of determining the affected individual's cognitive reserve (preserved premorbid abilities and residual cognitive strengths), areas of cognitive function that need further assessment, and cognitive deficits that may interfere with the treatment, the extent of change in cognitive function with respect to estimates of premorbid function, and functional ability in real-life settings. The Wechsler scales [Wechsler Adult Intelligence Scale III (WAIS-III), Wechsler Abbreviated Scale of Intelligence, and Wechsler Intelligence Scale for Children IV] are commonly used to assess intellectual ability following TBI.

Memory and Learning

Memory is one of the more vital cognitive functions. Memory deficits lead to dependency, isolation, and interruption of a sense of personal continuity.⁶ Memory impairments affect the individual's ability to learn and retain new information. Memory impairments post-TBI often interfere with cognitive rehabilitation as they affect attendance and compliance. Assessment of memory facilitates the identification of impairments in particular memory systems (eg, explicit/implicit and verbal/visual) and in the specific memory process that has been affected (eg, encoding, storage, or retrieval). A variety of measures are used to assess verbal memory (eg, California Verbal Learning Test, Hopkins Verbal Learning Test, and Rey Auditory Verbal Learning Test) and visual memory (Benton Visual Retention Test and Rey-Osterrieth Complex Figure Test). In addition, there are comprehensive memory batteries, such as the Wechsler Memory Scale, that comprehensively examine aspects of auditory and visual declarative and working memory.⁷

Psychomotor Function

Manual dexterity is used to examine changes in the differential superiority of the dominant hand

versus the nondominant hand as a means of providing a quick assessment of the diffuseness of the brain injury. Measures such as finger tapping and grip strength provide additional useful information.⁷ A handful of measures of motor function that are commonly included in a neuropsychological assessment following TBI include the Finger Tapping Test, Grooved Pegboard, and Purdue Pegboard.

Processing Speed

In addition to motor speed, the speed at which information is processed is assessed post-TBI. Following TBI, the rate of mental activity may be slowed, and this reduction in speed manifests itself as a delayed reaction time and/or increased task completion time. Several commonly used tests that measure processing speed are embedded in comprehensive batteries, such as the WAIS-III (Digit-Symbol Coding and Symbol Search subtests) and the Woodcock-Johnson Test of Cognitive Abilities (Decision Speed and Visual Match subtests). Additionally, such measures as the Stroop Color Word Test and the Trail Making Test are used to examine processing speed.

Attention

Attention is a complex mental activity that refers to how an individual receives and begins to process internal and external stimuli.⁶ Impairments in attention following TBI may range from difficulty remaining focused and ignoring distractions to failure to complete tasks involving multiple steps and learning new information. Several aspects of attention are assessed after TBI, including orientation, concentration, vigilance, distractibility, working memory, and multitasking. Commonly used measures of visual attention include the Conner's Continuous Performance Test and the Symbol Digit Modalities Test. Measures of auditory attention include the Paced Auditory Serial Attention Test (PASAT) and subtests of the WAIS-III (Digit Span and Letter-Number Sequencing), which assess auditory attention.

Language

Verbal communication following TBI may be impeded by deficits in receptive and expressive language. Individuals with TBI may have difficulty understanding conversation and instructions (oral or written), responding to requests for information, expressing their needs, and maintaining conversation. Language measures include comprehensive batteries

(eg, the Multilingual Aphasia Examination and the Boston Diagnostic Aphasia Examination) or individual tests of specific language functions, such as the Boston Naming Test to assess visual naming ability, the Peabody Picture Vocabulary Test to assess receptive vocabulary, and the Controlled Oral Word Association Test to assess verbal fluency.

Executive Function

Executive function refers to the mental capacity to "engage successfully in independent, purposive, self-serving behavior" (Lezak *et al.*,⁶ p 31). Executive function impairments post-TBI affect the individual's ability to initiate, plan, set goals, monitor performance, anticipate consequences, and respond flexibly and adaptively.⁷ Frequently, executive dysfunction is the most disabling of all cognitive impairments as difficulties in this domain are pervasive because they can affect all aspects of the person's ability to function effectively in his personal or professional life. There are several measures of executive function, including multitest batteries (eg, the Behavioral Assessment of the Dysexecutive Syndrome and the Delis-Kaplan Executive Function System) as well as single-test measures (eg, the Category Test, the Stroop Test, and the Wisconsin Card Sorting Test).

EVIDENCE FOR EFFECTIVENESS OF COGNITIVE REHABILITATION

The toolkit of cognitive rehabilitation interventions has grown rapidly in the last 3 decades. Several interventions for individuals with TBI have been developed to improve visual-perceptual skills, language, attention, memory, and executive functioning. As cognitive rehabilitation interventions began to emerge and proliferate, the need for evidence supporting their efficacy and effectiveness grew as well. A subcommittee of the Brain Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine conducted 2 systematic reviews of the existing literature on cognitive rehabilitation interventions for TBI and stroke in order to assess the status of the evidence and to provide recommendations for clinical practice.^{3,4} An extensive search using several keywords, including attention, awareness, cognition, communication, executive, language, memory, perception, problem solving, reasoning, rehabilitation, remediation, and training, was conducted. The studies that were identified were subsequently classified by the cognitive skill that was targeted for treatment: attention,

visual perception, apraxia, language and communication, memory, executive functioning, problem solving and awareness, and comprehensive-holistic cognitive rehabilitation. The first review included 171 intervention studies published between 1971 and 1997.³ Of these, 29 were rated as class I (prospective, randomized or quasi-randomized controlled studies), 35 were rated as class II (prospective, nonrandomized cohort or retrospective, nonrandomized case-control or clinical series with controls), and 107 were rated as class III (clinical series without controls or single-case designs). The second review included 87 studies published between 1998 and 2002; 17 were rated as class I, 8 were rated as class II, and 62 were rated as class III.⁴ The same special interest group is currently reviewing cognitive rehabilitation studies published between 2003 and 2007.

The conclusions drawn from the 2 reviews supported the effectiveness of cognitive rehabilitation interventions. The authors reported that the cognitive rehabilitation interventions that they reviewed were more effective than traditional types of treatment.^{3,4} Cicerone *et al.*^{3,4} used these reviews to develop evidence-based practice standards, guidelines, and options for practitioners working with individuals with TBI in the domains of attention, memory, and executive function. Practice standards were based on solid evidence from class I or class II studies and affirmed the efficacy of cognitive rehabilitation interventions in the form of strategy training to improve attention and memory and interventions to improve functional communication and conversational skills. Practice guidelines were based on class II evidence and included recommendations for attention training (varying modalities, levels of complexity, and response demands), methods to improve reading comprehension and language formation, training in problem-solving strategies, and participation in comprehensive neuropsychological rehabilitation. Finally, practice options were based on class II and class III evidence and included the use of memory notebooks or other external aids to address memory impairments and the use of methods of self-instruction, self-questioning, and self-monitoring to improve executive function. In addition, Cicerone *et al.*⁴ argued that the next step for cognitive rehabilitation research is to investigate the relationship between treatment outcome and patient characteristics.

COGNITIVE REHABILITATION INTERVENTIONS AFTER TRAUMATIC BRAIN INJURY

Cognitive rehabilitation interventions are applicable at all stages of postinjury recovery (acute, subacute, and postacute) and in different settings (eg, inpatient, outpatient, and domestic environments). In addition, they can be administered in different modalities (eg, individual, family, and group) and by professionals in different disciplines (eg, neuropsychologists, occupational therapists, and speech-language pathologists) with proper training and supervision. The following sections include an overview of selective evidence-based cognitive rehabilitation interventions that address impairments in attention, memory, and executive functions following TBI.

Cognitive Rehabilitation for Attention

Interventions for attention deficits range from simple tasks such as using flashcards to improve basic attention skills⁸ to more complex tasks to improve complex attention and working memory using a variety of visual and verbal tasks.^{9–11} Sohlberg and Mateer¹² developed a training program to improve visual and auditory attention. The Attention Process Training (APT) program (APT-I and APT-II) is a theory-based intervention that hierarchically organizes attention into 5 components: focused attention, sustained attention, selective attention, alternating attention, and divided attention. The training program consists of various tasks organized by increasing difficulty. Competence is achieved through repetition at a certain skill level and progression to a higher skill level when the easier task is mastered. In addition, the developers have incorporated generalization tasks within the intervention to increase generalizability. Treatment is provided through a combination of individual and group treatments by trained clinicians. There is ample evidence supporting the effectiveness of APT.^{13–15} The ease of implementing the treatment notwithstanding, clinical skill is a necessary component of any intervention. The role of the clinician is indispensable, and the APT is a good tool only insofar as it is applied by therapists with a good understanding of the training program as well as the deficits of the person being treated.

Other interventions for attention have included using the PASAT or variations of the PASAT such as the n-back procedure. PASAT-type tasks require listening to a sequence of stimuli and responding according to a rule. For example, in the original

PASAT, respondents are required to listen to a sequence of numbers and utter aloud the sum of each number when it is added to the number that immediately preceded it. Evidence suggests that use of these interventions leads to improvements in measures of working memory and self-report of attention difficulties.^{9,11}

Cognitive Rehabilitation for Memory

Interventions for improving memory are an essential component of cognitive rehabilitation. Several interventions focus on improving different aspects of memory, such as face-name associations, memory for past events, prospective memory, and learning new information. Memory interventions have included restorative approaches, such as word-list learning, paragraph listening, visual imagery, and use of mnemonic strategies,^{16,17} as well as compensatory approaches, such as the use of memory notebooks and other methods of recording and tracking information and the use of assistive technology tools, including personal computers, portable electronic devices, voice recorders, and pagers, which have been found to be instrumental in enhancing prospective memory (ie, the ability to remember to carry out a certain action at a specified time in the future or in response to a specific future event).^{18–21}

One of the most effective interventions for the rehabilitation of memory impairments post-TBI is the memory notebook. The use of a memory notebook has been supported empirically.^{22,23} Sohlberg and Mateer¹⁶ presented a systematic method for using a memory notebook to facilitate learning. A memory notebook usually includes sections for orientation (autobiographical and/or injury-related information), a memory log, a calendar, to-do lists, transportation (eg, maps, public transportation schedules, and taxi phone numbers), a feelings log (a section to record emotions occurring in certain situations), names (names and other identifying information about other individuals), and any other section that may be personally relevant. There are 3 phases to learning how to use a memory book: acquisition, application, and adaptation. During the acquisition phase, the individual with TBI becomes familiar with the different sections and the purpose of the notebook. During the application phase, the individual learns to use the memory notebook in simulated settings. Finally, during the adaptation phase, the person learns to extend the use of the memory notebook to naturalistic environments, such as household or vocational settings. There are several factors that interfere with the successful use of memory notebooks and other external memory aids. They

include client characteristics (eg, a lack of awareness of the need for the memory notebook and an unwillingness to draw attention to oneself), features of the memory notebook per se (eg, too simple or too complicated), environmental factors (eg, a low level of demand, especially if the patient is home-bound), and clinician variables (eg, training approach and identification of client-relevant sections).²⁴

Cognitive Rehabilitation for Executive Functioning

Interventions to improve executive functioning following TBI have targeted problem solving,^{25–28} planning and organization,^{29,30} goal-directed behavior,^{31,32} and self-monitoring and self-regulation.^{25,33,34} In a meta-analysis of intervention studies for executive functioning post-TBI, Kennedy *et al.*³⁵ reported favorable findings in terms of improvements in problem solving, planning, organization, and multitasking. Improvements extended to personally relevant functional activities, and treatment effects were maintained and generalized. The authors concluded that interventions using metacognitive strategies appeared to be the most efficacious and effective. Metacognitive strategy instruction was the type of intervention used across several studies that were reviewed, including randomized clinical trials and single-case studies. Metacognitive strategy instruction includes using and internalizing step-by-step procedures intended to enhance problem solving, planning, organization, and multitasking by increasing the capacity for self-regulation.³⁵ Metacognitive strategies aim at improving self-regulation by increasing self-awareness, which promotes the formation of personally relevant goals, self-monitoring, which enables individuals to assess their performance and reduce or prevent errors, and self-control, which facilitates initiation and behavioral change.³⁵ Interventions for executive functioning are administered individually, in groups, or with a combination of individual and group treatments.

Comprehensive-Holistic Day Treatment Programs

Comprehensive-holistic day treatment programs (CHPs) offer a combination of therapeutic services and are a popular treatment option for individuals with TBI. A typical CHP includes individual and group cognitive rehabilitation, psychotherapy, psycho-education, and family therapy. Interventions are focused on specific cognitive domains, and group interventions are used to focus on executive dysfunction and to promote generalization of learning.

Within such programs, treatment goals are articulated clearly, and progress is monitored regularly. The goals of treatment include improving cognitive function, increasing awareness, and addressing interpersonal, social, and emotional concerns. Treatment is usually administered by a transdisciplinary team, it incorporates community activities and vocational trials, and involvement of significant others is highly encouraged.³⁶ The evidence supports the effectiveness of CHPs, which in addition to resulting in improvements in cognitive functioning appear to facilitate skill transfer and generalization and to increase self-awareness, behavioral and affective regulation, psychosocial functioning, and community integration.^{3,4} Comprehensive-holistic approaches to treatment have thus become the standard of care in rehabilitation following TBI.³⁷

Two randomized clinical trials are currently in progress at the Mount Sinai School of Medicine to examine the efficacy of a CHP consisting of individual and group interventions for executive dysfunction.²⁵ The intervention is based on a metacognitive strategy intended to improve problem-solving and self-regulation abilities. The intervention entails the following:

- Individual training to improve diverse aspects of attention (selective, sustained, alternating, and divided).
- Group and individual instruction in a step-by-step problem-solving procedure that facilitates identification of problems, awareness of various aspects of problems, generation of alternatives, initiation of action, and self-monitoring.
- Group and individual training in strategies to improve emotional regulation. Individualized goal setting, repetition, and feedback are used extensively in order to enhance maintenance of the self-instructional abilities and to achieve generalization of the use of the metacognitive strategies to a variety of real-life settings.

Individuals receive the experimental CHP treatment, a standard-of-care CHP treatment, or no treatment (wait-list control group). The efficacy of the experimental intervention will be based on differences in performance on standardized neuropsychological tests and psychosocial measures. It is expected that individuals with TBI receiving the experimental intervention will exhibit superior performance on neuropsychological measures and improved psychosocial outcome in comparison with the wait-list control group or the standard-of-care treatment group.

Use of Technology in Cognitive Rehabilitation

The use of technology in neuropsychological rehabilitation ranges from using computers as passive tools to project visual and verbal stimuli during cognitive training to using technological aids actively and in naturalistic settings as compensatory tools or cognitive orthotics.³⁸ Computer use in cognitive rehabilitation extends to memory training,³⁹⁻⁴¹ attention,^{39,42} problem solving,³⁹ and job simulation.⁴³ Technological aids are frequently used in cognitive rehabilitation to improve performance on cognitive tasks. For example, a recent randomized control trial showed that use of a pager has been found to enhance prospective memory, which refers to the ability to remember to perform an activity in the future.¹⁹ In an extensive review of existing technological aids, LoPresti *et al.*³⁸ classified aids into technologies for memory and executive function impairments and technologies for information processing impairments. Devices for memory and executive function compensation include digital watches, alarms, voice organizers, mobile phone-computer interactive systems, and handheld devices, such as personal digital assistants. Devices for information processing impairments include a keyboard for typing instead of writing, software that alters the features of text (eg, size and color) on a computer screen to facilitate reading, and speech output/speech recognition software.

Although the use of technology in TBI rehabilitation permits the administration of tasks that would be otherwise difficult to administer (eg, tracking reaction time in milliseconds or using a pager to cue oneself), these aids are intended not to replace a therapist but to augment the therapeutic experience. Indeed, research has failed to show an advantage of computer-assisted interventions over traditional interventions,^{39,42} once again reminding us that the clinician is an active ingredient in the treatment. The role of the therapist in cognitive rehabilitation is "to set and maintain the structure, determine treatment needs and readiness, provide feedback and guidance, teach and reinforce the use of tools, and process emotional reactions,"⁴⁴ which to date no technological aid can accomplish successfully.

ROLE OF TIME SINCE INJURY IN COGNITIVE REHABILITATION

There is no empirical evidence supporting the idea that there is a critical period during which cognitive rehabilitation is more effective.⁴⁵ Research on the role of time since injury in treatment is divided,

Table 1. *Key Concepts in Cognitive Rehabilitation Following Traumatic Brain Injury.*

Cognitive rehabilitation is a treatment for cognitive impairments related to traumatic brain injury that is strongly supported by well-designed research.
A neuropsychological assessment is required in order to assess cognitive function and develop an appropriate treatment plan.
Cognitive rehabilitation consists of diverse interventions that must be tailored to the individual patient.
Cognitive rehabilitation can be effective regardless of the length of time since the injury and the injury severity level.
Cognitive rehabilitation leads to improvements in cognitive and psychosocial functioning.

with some studies showing no difference in terms of the benefits of treatment in neuropsychological measures and psychosocial and vocational outcomes and other studies showing only a slight advantage of early rehabilitation.⁴⁴ Recent studies about brain plasticity suggest that interventions delivered at later stages post-injury may be more beneficial.⁴⁶ The lack of strong evidence about the optimal timing of the intervention suggests that there is no critical window for new learning. Learning can occur at any point post-injury, and there are no time limits in terms of the benefits that can accrue to the person from cognitive rehabilitation. Making individuals with TBI aware that treatment is beneficial at any time post-injury is likely to instill hope and increase their motivation and eventually lead to improvements in cognitive, emotional, and psychosocial functioning.

ROLE OF INJURY SEVERITY IN COGNITIVE REHABILITATION

Cognitive rehabilitation interventions are used in the treatment of individuals with TBI whose severity of injury ranges from mild to severe. Intervention studies typically report the level of severity of the participants using traditional measures of injury severity (eg, the Glasgow Coma Scale or the duration of posttraumatic amnesia). Little research, however, exists that directly examines the role of injury severity in cognitive rehabilitation.⁴⁵ The literature generally suggests that severity of injury per se may not be as instrumental in designing and delivering a cognitive rehabilitation intervention as is identifying particular cognitive strengths and weaknesses and functional deficits of the person.⁴⁶ The target, intensity, and difficulty level of the intervention are better determined by the individual's cognitive profile, functional ability, level of self-awareness, and availability of environmental supports than by severity of injury.

CONCLUSION

Cognitive rehabilitation interventions have proliferated in the last 30 years and have been implemented

widely in the treatment of the cognitive deficits resulting from TBI (Table 1TBL 1). The toolkit includes interventions that can be adapted to facilitate improvements in functioning at different levels of severity, and that is not limited by the length of time since injury. Evidence from class I and class II studies has been amassed that supports the effectiveness of these interventions not only in terms of improving cognitive functioning but also in terms of improved psychosocial functioning and vocational outcomes. In order to render these interventions more successful, "they must be embedded in an appropriate context, be delivered systematically and creatively, and be individualized to fit the unique cognitive and psychotherapeutic needs of each individual" (Gordon and Hibbard,⁴⁵ p 660). Further research is still needed to identify the patient and treatment factors that contribute to successful outcome, to explicate the theoretical models underlying the interventions, and to identify the extent of the clinical significance of these interventions. Cognitive rehabilitation interventions are promising treatments that contribute to the well-being and quality of life of individuals with TBI.

ACKNOWLEDGMENT

The preparation of this article was supported in part by grants from the US Department of Education National Institute on Rehabilitation Research (H133B040033 and H133A070033) and from the Centers for Disease Control and Prevention (1R49CE001171-01).

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Ashman TA, Gordon WA, Cantor JB, Hibbard MR. Neurobehavioral consequences of traumatic brain injury. *Mt Sinai J Med* 2006; 73: 999–1005.

2. Cicerone KD, Mott T, Azulay J, Friel JC. Community integration and satisfaction with functioning after intensive cognitive rehabilitation for traumatic brain injury. *Arch Phys Med Rehabil* 2004; 85: 943–950.
3. Cicerone KD, Dahlberg C, Kalmar K, et al. Evidence-based cognitive rehabilitation: recommendations for clinical practice. *Arch Phys Med Rehabil* 2000; 81: 1596–1615.
4. Cicerone KD, Dahlberg C, Malec JF, et al. Evidence-based cognitive rehabilitation: updated review of the literature from 1998 through 2002. *Arch Phys Med Rehabil* 2005; 86: 1681–1692.
5. Gordon WA, Zafonte R, Cicerone K, et al. Traumatic brain injury rehabilitation: state of the science. *Am J Phys Med Rehabil* 2006; 85: 343–382.
6. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. 4th ed. New York, NY: Oxford University Press; 2004.
7. Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3rd ed. New York, NY: Oxford University Press; 2006.
8. Zencius AH, Wesolowski MD, Rodriguez IM. Improving orientation in head injured adults by repeated practice, multi-sensory input and peer participation. *Brain Inj* 1998; 12: 53–61.
9. Cicerone KD. Remediation of “working attention” in mild traumatic brain injury. *Brain Inj* 2002; 16: 185–195.
10. Parente R, Kolakowsky-Hayner S, Krug K, Wilk C. Retraining working memory after traumatic brain injury. *NeuroRehabilitation* 1999; 13: 157–163.
11. Serino A, Ciaramelli E, Santantonio AD, et al. A pilot study for rehabilitation of central executive deficits after traumatic brain injury. *Brain Inj* 2007; 21: 11–19.
12. Sohlberg MM, Mateer CA. Effectiveness of an attention-training program. *J Clin Exp Neuropsychol* 1987; 9: 117–130.
13. Park NW, Ingles JL. Effectiveness of attention rehabilitation after an acquired brain injury: a meta-analysis. *Neuropsychology* 2001; 15: 199–210.
14. Palmese CA, Raskin SA. The rehabilitation of attention in individuals with mild traumatic brain injury, using the APT-II programme. *Brain Inj* 2000; 14: 535–548.
15. Pero S, Incoccia C, Caracciolo B, et al. Rehabilitation of attention in two patients with traumatic brain injury by means of ‘attention process training’. *Brain Inj* 2006; 20: 1207–1219.
16. Sohlberg MM, Mateer CA. Training use of compensatory memory books: a three stage behavioral approach. *J Clin Exp Neuropsychol* 1989; 11: 871–891.
17. Sohlberg MM, White O, Evans E, Mateer C. Background and initial case studies into the effects of prospective memory training. *Brain Inj* 1992; 6: 129–138.
18. van den Broek MD, Downes J, Johnson Z, et al. Evaluation of an electronic memory aid in the neuropsychological rehabilitation of prospective memory deficits. *Brain Inj* 2000; 14: 455–462.
19. Wilson BA, Emslie H, Quirk K, et al. A randomized control trial to evaluate a paging system for people with traumatic brain injury. *Brain Inj* 2005; 19: 891–894.
20. Hart T, Hawkey K, Whyte J. Use of a portable voice organizer to remember therapy goals in traumatic brain injury rehabilitation: a within-subjects trial. *J Head Trauma Rehabil* 2002; 17: 556–570.
21. Wright P, Rogers N, Hall C, et al. Comparison of pocket-computer memory aids for people with brain injury. *Brain Inj* 2001; 15: 787–800.
22. Schmitter-Edgecombe M, Fahy JF, Whelan JP, Long CJ. Memory remediation after severe closed head injury: notebook training versus supportive therapy. *J Consult Clin Psychol* 1995; 63: 484–489.
23. Zencius A, Wesolowski MD, Krankowski T, Burke WH. Memory notebook training with traumatically brain-injured clients. *Brain Inj* 1991; 5: 321–325.
24. McKerracher G, Powell T, Oyeboode J. A single case experimental design comparing two memory notebook formats for a man with memory problems caused by traumatic brain injury. *Neuropsychol Rehabil* 2005; 15: 115–128.
25. Gordon WA, Cantor J, Ashman T, Brown M. Treatment of post-TBI executive dysfunction: application of theory to clinical practice. *J Head Trauma Rehabil* 2006; 21: 156–167.
26. Rath JF, Simon D, Langenbahn DM, et al. Group treatment of problem-solving deficits in outpatients with traumatic brain injury: a randomised outcome study. *Neuropsychol Rehabil* 2003; 13: 461–488.
27. Foxx RM, Martella RC, Marchand Martella NE. The acquisition, maintenance, and generalization of problem solving skills by closed head injured adults. *Behav Ther* 1989; 20: 61–76.
28. von Cramon DY, Matthes von Cramon G, Mai N. Problem solving deficits in brain injured patients: a therapeutic approach. *Neuropsychol Rehabil* 1991; 1: 45–64.
29. Cicerone KD, Wood JC. Planning disorder after closed head injury: a case study. *Arch Phys Med Rehabil* 1987; 68: 111–115.
30. Fish J, Evans JJ, Nimmo M, et al. Rehabilitation of executive dysfunction following brain injury: “content-free” cueing improves everyday prospective memory performance. *Neuropsychologia* 2007; 45: 1318–1330.
31. Levine B, Robertson IH, Clare L, et al. Rehabilitation of executive functioning: an experimental-clinical validation of goal management training. *J Int Neuropsychol Soc* 2000; 6: 299–312.
32. Manly T, Hawkins K, Evans J, et al. Rehabilitation of executive function: facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia* 2002; 40: 271–281.
33. Medd J, Tate RL. Evaluation of an anger management therapy programme following acquired brain injury: a preliminary study. *Neuropsychol Rehabil* 2000; 10: 185–201.
34. Kennedy MR, Coelho C. Self-regulation after traumatic brain injury: a framework for intervention of memory and problem solving. *Semin Speech Lang* 2005; 26: 242–255.
35. Kennedy MR, Coelho C, Turkstra L, et al. Intervention for executive functions after traumatic brain injury: a systematic review, meta-analysis and clinical recommendations. *Neuropsychol Rehabil* 2008; 18: 257–299.
36. Malec JF, Basford JS. Postacute brain injury rehabilitation. *Arch Phys Med Rehabil* 1996; 77: 198–207.
37. Prigatano GP. *Principles of Neuropsychological Rehabilitation*. New York, NY: Oxford University Press; 1999.

38. LoPresti EF, Mihailidis A, Kirsch N. Assistive technology for cognitive rehabilitation: state of the art. *Neuropsychol Rehabil* 2004; 14: 5–39.
39. Chen SHA, Thomas JD, Glueckauf RL, Bracy OL. The effectiveness of computer-assisted cognitive rehabilitation for persons with traumatic brain injury. *Brain Inj* 1997; 11: 197–209.
40. Dou ZL, Man DWK, Ou HN, et al. Computerized errorless learning-based memory rehabilitation for Chinese patients with brain injury: a preliminary quasi-experimental clinical design study. *Brain Inj* 2006; 20: 219–225.
41. Goldstein G, Beers SR, Shemansky WJ, Longmore S. An assistive device for persons with severe amnesia. *J Rehabil Res Dev* 1998; 35: 238–244.
42. Batchelor J, Shores EA, Marosszeky JE, et al. Cognitive rehabilitation of severely closed-head injured patients using computer assisted and noncomputerized treatment techniques. *J Head Trauma Rehabil* 1988; 3: 78–85.
43. Kirsch NL, Levine SP, Lajiness-O'Neill R, Schnyder M. Computer-assisted interactive task guidance: facilitating the performance of a simulated vocational task. *J Head Trauma Rehabil* 1992; 7: 13–25.
44. Tsaousides T, Gordon WA. Neuropsychological interventions following traumatic brain injury. In: Ashley MJ, ed. *Traumatic Brain Injury: Rehabilitative Treatment and Case Management*. 3rd ed. Boca Raton, FL: Taylor & Francis. In press.
45. Gordon WA, Hibbard MR. Cognitive rehabilitation. In: Silver JM, McAllister TW, Yudofsky SC, eds. *Textbook of Traumatic Brain Injury*. Washington, DC: American Psychiatric Publishing; 2005; 655–660.
46. Kennedy MR, Turkstra L. Group intervention studies in the cognitive rehabilitation of individuals with traumatic brain injury: challenges faced by researchers. *Neuropsychol Rev* 2006; 16: 151–159.

Neurorehabilitation in Traumatic Brain Injury: Does It Make a Difference?

Brian D. Greenwald, MD,¹ and John L. Rigg, MD²

¹Department of Rehabilitation Medicine, Mount Sinai Medical Center, New York, NY

²Traumatic Brain Injury Program, Dwight D. Eisenhower Army Medical Center, Southeast Regional Medical Command, Fort Gordon, GA

ABSTRACT

Each year, 1.4 million people in the United States are seen in a hospital for a traumatic brain injury. Those with moderate-to-severe traumatic brain injury frequently go through a course of inpatient neurorehabilitation prior to discharge back into the community. A broad overview of neurorehabilitation is presented, including the standards for admission to inpatient rehabilitation and the members and roles of the neurorehabilitation team. Common medical complications that are managed after moderate-to-severe traumatic brain injury are reviewed. The spectrum of arousal issues is summarized. The evidence regarding neurorehabilitation is then reviewed. Future studies that are underway to better understand the utility of neurorehabilitation are then discussed. *Mt Sinai J Med* 76:182–189, 2009. © 2009 Mount Sinai School of Medicine

Key Words: brain injury, evidence, rehabilitation.

The Centers for Disease Control and Prevention (CDC) estimates that each year 1.4 million Americans are seen in a hospital for a traumatic brain injury (TBI).¹ Of these patients, 50,000 die, 235,000 are hospitalized, and 1.1 million are treated and released

from an emergency department. The number of people who sustain a TBI and do not seek medical care is unknown. Sports-related concussions are routinely underreported, and it has been approximated that there are 1.6 million to 3.8 million cases each year.¹

It is estimated that 80,000 people in the United States are left with long-term disabilities from TBI annually.¹ The CDC estimates that at least 5.3 million Americans, approximately 2% of the US population, currently have a long-term or lifelong need for help to perform activities of daily living as a result of a TBI.¹

Severity of TBI is most commonly measured with the Glasgow Coma Scale. Mild brain injury represents 80% of all TBI, moderate brain injury represents 10%, and severe brain injury represents 10%.² As a broad generalization, the higher the Glasgow Coma Scale score is on admission to the emergency department, the more likely a good outcome will result. Despite this, estimates are that 15% of those with mild brain injury will have long-term symptoms.

Patients with moderate-to-severe TBI are most likely to be hospitalized and considered for inpatient rehabilitation. Because of the traumatic nature of their injury, they often have multiple comorbid injuries. In addition, the elderly frequently have premorbid conditions that complicate their injury and recovery.

The spectrum of disability varies widely after moderate-to-severe TBI, ranging from cognitive impairment with no physical impairment to deep coma. Admission criteria to specialized brain injury rehabilitation centers also vary. Table 1 presents criteria for admission to inpatient rehabilitation. The decision of whether a patient would be best served in an acute (intensive) or subacute rehabilitation center is generally based on the amount of hours that he can participate in therapy per day. To qualify for acute rehabilitation, the patient should be able to participate in at least 3 hours of 2 or more therapies per day 5 to 7 days per week. Subacute rehabilitation

Address Correspondence to:

Brian D. Greenwald, MD

Assistant Professor

Department of Rehabilitation
Medicine

Mount Sinai School of Medicine
New York, NY

Email: brian.greenwald@mountsinai.org

Table 1. *Criteria for Inpatient Rehabilitation.*

The patient has an acute disability that prevents him or her from returning home with family care.
Medical or surgical conditions are sufficiently stable to allow participation in therapies.
The patient demonstrate the ability to participate in at least 1 hour of therapy 2 times a day.
The patient demonstrates the ability to make progress in acute care therapies.
The patient has a social support system that will allow him or her to return home after reasonable improvement of function.

is often done in specialized units in nursing homes (skilled nursing facilities). This is more appropriate for patients who can participate in only 0.5 to 2 hours of therapy per day. In addition to these criteria, other factors are often considered in rehabilitation placement, such as the need for medical supervision, age, likelihood of return to home, and payor source. Neurorehabilitation is a young field with an evolving body of literature supporting its impact on outcomes. This article provides an overview of the field and directions for future research.

METHODOLOGY

Articles were searched with PubMed. Search terms included *neurorehabilitation*, *rehabilitation*, and *brain injury*. Search terms were then combined, and articles were chosen that were directly related to the evidence for neurorehabilitation in TBI patients. References were obtained from the CDC Web site for epidemiology. References for medical complications of TBI were obtained through a PubMed search of TBI and complications.

NEUROREHABILITATION

An injury to the brain will likely result in a multitude of functional problems that are impossible to predict on the basis of traditional diagnostic studies alone. Head computed tomography and magnetic resonance imaging may be negative, and yet a patient may exhibit obvious clinical signs of functional disruption. Damage to the parenchyma of the brain can occur on the microscopic level, not resulting in tissue alteration that can be seen in traditional diagnostic studies. A comprehensive clinical assessment of a patient's functional deficits must be made in order to determine an effective course of treatment.

Table 2. *Disabilities That May Arise as a Result of Traumatic Brain Injury.*

Cognitive processing	Memory
	Concentration
	Judgment
	Problem solving
Motor skills	Strength
	Coordination
	Ambulation
Sensory perception	Smell
	Vision
	Hearing
	Tactile sensation
Mood	Instability
	Impulsivity
	Irritability

Disabilities that may arise as a result of a TBI commonly include problems with (1) cognitive processing, including memory, concentration, processing speed, and judgment; (2) motor skills, including strength, coordination, and ambulation; (3) sensory problems, including special senses such as smell, vision, and hearing as well as tactile sensation; and (4) emotional/mood issues, including instability, impulsivity, and irritability (see Table 2). It is important to remember that the TBI population is not a cross section of society. Those with behavioral, psychiatric, and learning disabilities are disproportionately represented, and the premorbid personality may be magnified after a brain injury.³

The primary goals of a rehabilitation program are to maximize function, minimize impairments, and prevent complications. A TBI will typically result in a constellation of impairments that demands a multidisciplinary team approach in order to attain those goals and maximize functional recovery. Return of function is not limited to strictly physical function but also includes emotional recovery and social/community reintegration, including vocational rehabilitation with a goal of returning to work. Table 3 presents the disciplines that contribute to

Table 3. *Traumatic Brain Injury Rehabilitation Team.*

Physiatrist
Rehabilitation nurse
Physical therapist
Occupational therapist
Speech/language pathologist
Neuropsychologist
Nutritionist/dietician
Social worker
Case manager
Recreational therapists
Vocation therapists

the rehabilitation team. Each of these disciplines provides a highly trained practitioner to maximize a patient's functional recovery. The members of the interdisciplinary team will typically have undergone specific training to assist the patient in dealing with the unique aspects of brain injury rehabilitation. A physiatrist generally functions as the leader of the team, coordinating the multiple aspects of a brain-injured patient's care. He is responsible for identifying rehabilitation needs and managing medical complications.

The physiatrist is responsible for the medical management of neuropsychiatric symptoms, sleep, arousal, pain management, and nutrition. An in-depth understanding of preventing, diagnosing, and treating medical problems associated with TBI is a critical element of this specialty. Common medical complications seen after TBI are discussed later. Prescribing tailored therapy orders for speech/language pathology, physical therapy, and occupational therapy guides the treatment while the patient is in rehabilitation. Reassessing and updating these orders is needed as the patient improves. Family and patient education play a critical role in acclimating them to this new disability.

The clinical psychologist will meet with the patient to determine if there are any psychological issues that are preventing that patient from fully participating in, and benefiting from, his therapy program. He may also provide counseling and education to the patient's family members, thus ensuring that they have an understanding of the treatment plan and possible outcomes.

The neuropsychologist will perform extensive testing to evaluate cognition, mood, personality, and behavior. The testing will quantify a patient's higher cortical functioning and reveal aspects of attention, memory, speed of information processing, language, visual-spatial ability, sensory processing, motor ability, higher order executive functioning, and intelligence. Cognitive remediation on the inpatient basis will vary from group to individual treatments, depending on the patient's needs.

The case manager is responsible for coordinating the patient's care and is also involved in family training/counseling. The case manager is also responsible for setting up the details of the discharge plan, including follow-up appointments with physicians and physical, occupational, and speech therapists, and ensuring the availability of any special equipment that the patient may need.

The physical therapist's role is to treat any physical dysfunction or injury, particularly with respect to ambulation. This is accomplished via the use of

therapeutic exercise and the application of modalities (eg, heat, ice, and electrical stimulation) with a goal of restoring normal function. Specific goals to be accomplished in the physical therapy gym include strengthening and improving endurance, ambulation, and balance.

Occupational therapists provide training in real-life activities that are normally part of a patient's routine function, which are called activities of daily living, in order to maximize a patient's independence. These will typically include eating, bathing, grooming, dressing, and functional transfers, including bed to wheelchair, toilet, tub, and shower transfers.

The speech-language pathology therapist is responsible for the treatment of speech, swallow, communication, and cognitive disorders. The therapeutic regimen may include physical exercises to strengthen the muscles used in speech/swallowing, speech drills to improve clarity, and sound production practice to improve articulation. Specially trained speech therapists may perform VitalStim therapy, which consists of electrical stimulation applied to the throat to improve the swallow function of patients who are unable to eat because of their inability to swallow. A speech-language pathology therapist will also assess a patient's cognitive status and initiate cognitive training programs. An example of this would be a daily orientation group, which can help brain-injured patients tune into the calendar as well as local weather and news, sports, and entertainment events.

Often, particularly in the case of severe TBI, patients cannot return to their previous employment because of disabilities. Vocational counselors acquire and evaluate information about the brain-injured patient's functional skills and perform testing and assessments of the patient to determine potential careers for which the patient may be suitable at discharge. Testing to determine jobs in which the patient may be interested and also that he may be capable of doing successfully is suggested. Job training can then be set up as part of the rehabilitation process.

The role of the recreational therapist is to provide recreational resources and opportunities in order to improve health and well-being. This may include outings (eg, to community parks, athletic events, and museums) or in-hospital social and group activities (eg, games and arts and crafts).

An important part of the comprehensive rehabilitation program is the team conference. It is a meeting of all the individuals who are involved in a patient's care and typically is held once a week. Each member of the rehabilitation team has an opportunity to present the details of a patient's functional progress in his particular discipline, thus providing all of the team members a better understanding of

how a patient may be progressing. For example, a patient may report that he is doing fine to the physician during daily rounds. However, the physical therapist who spends an hour a day with the patient doing exercises in the gym may report that the patient's progress in therapy has been slow or that his participation has been poor. The patient's level of participation can be discussed, and a determination of what is slowing down his recovery can be made. For example, there are a number of factors that may be involved, such as fatigue, pain, abulia or other psychological issues, spasticity, or weakness. Once the reason that the patient has not been participating fully has been determined, appropriate treatment strategies can be determined to maximize his participation and recovery.

Inpatient rehabilitation is often only the first step in the rehabilitation process. After goals for a safe discharge are achieved in the inpatient rehabilitation setting, patients are generally transitioned to an outpatient brain injury rehabilitation program. Depending on the patient's needs, an outpatient program may look similar in therapy types to the inpatient program or may be more focused. Long-term cognitive and behavioral impairments are the most frequent complaints from patients and their families. Ponsford *et al.*⁴ interviewed 175 TBI patients 2 years after they had undergone acute rehabilitation. Most had sustained severe TBIs, and over 75% reported persisting cognitive, behavioral, or emotional problems as follows: memory problems (74%), fatigue (72%), word-finding difficulties (67%), irritability (67%), decreased speed of thinking (64%), and impaired concentration (62%).

COMMON MEDICAL COMPLICATIONS AFTER TRAUMATIC BRAIN INJURY

Common medical complications seen after TBI are listed in Table 4. Their etiology is linked directly to the brain injury and/or to extracranial injuries or comorbid conditions. Complications can occur early or late (more than 1 year) post-injury. Knowledge of signs and symptoms of these complications as well as treatment is important for the brain injury rehabilitation specialist but should also be familiar to the primary care specialist.

Posttraumatic Seizures

The overall risk of posttraumatic seizures varies widely, depending on the severity and mechanism of injury. Rates have been described as high as 53% after

Table 4. *Common Medical Complications of Traumatic Brain Injury.*

Posttraumatic seizures
Heterotopic ossification
Hydrocephalus
Endocrine
Fractures
Headaches
Peripheral nerve injuries
Arrhythmias
Dysphagia
Visual impairment
Spasticity
Dysautonomia
Pressure sores

military penetrating head injury and as low as <5% after mild brain injury with no imageable injury.⁵ It is generally accepted that seizure prophylaxis is given only for 7 days after moderate-to-severe TBI and then discontinued unless the patient has recurrent seizures.⁶ Long-term seizure prophylaxis is likely to have unwanted side effects and has not been shown to decrease risk of seizure after injury. No specific risk factors have been shown to guide which patients need long-term seizure prophylaxis.⁷

Heterotopic Ossification

Heterotopic ossification (HO) is a process in which new bone forms within soft tissues where bone formation does not usually occur. Clinical features include a warm, swollen, painful joint with restriction of range of movement. The pathophysiology of HO is not known, although it is thought to arise from a metaplasia of fibroblasts. HO is also seen after burns, joint replacement, and spinal cord injury. The areas that are most commonly involved after TBI are the soft tissues around the hip, elbow, shoulder, and knee. The incidence of HO in TBI patients has been reported to range from 11% to 77%, but the disease reaches clinical significance in only 11% to 35% of this group.⁸ Patients who sustain skeletal trauma, spasticity, immobilization, and coma for greater than 2 weeks are considered to be at highest risk for HO.⁸ HO is often quite painful, at times limiting a patient's ability to participate in rehabilitation. Treatments commonly used include range of motion, anti-inflammatories, bisphosphonates, and surgical resection.

Deep Vein Thrombosis

After moderate-to-severe TBI, all the commonly recognized risk factors for deep vein thrombosis

(DVT) are seen: stasis, hypercoagulable state, and endothelial damage (Virchow's triad). The reported incidence of DVT among patients with TBI is between 20% and 54%.^{9,10} Early chemical prophylaxis to prevent DVT in the immediate postinjury phase is fraught with the concern of initiating intracerebral hemorrhage. Despite the critical importance of this problem, there is a dearth of literature examining DVT prophylaxis risk/benefit in this population. At a minimum, early initiation of mechanical DVT prophylaxis is recommended, and venous ultrasound should be considered on admission to rehabilitation.

Spasticity

This common and potentially disabling result of TBI has options ranging from conservative (range of motion and splinting) to pharmacological (eg, baclofen, tizanidine, and dantrolene) to surgical (intrathecal baclofen pump and tendon lengthening). Combinations of treatments are often used to maximize efficacy. The time post-injury, severity of spasticity, number of limbs involved, and patient preference all need to be considered when spasticity treatment is tailored for an individual.

Endocrine

Over the last 5 years, there has been a growing awareness of the incidence of hypopituitarism after TBI. Recent literature suggests that 20% to 30% of patients after moderate-to-severe TBI may have pituitary dysfunction.¹¹ Signs and symptoms of hypopituitarism may mimic the sequelae of TBI. Screening for adrenal, thyroid, gonadal, and growth hormone deficiencies has been recommended at 3 and 12 months and earlier in some cases.¹¹ Outcome studies of large-scale screening and treatment are pending.

Spectrum of Arousal Impairments After Traumatic Brain Injury

A patient with a severe TBI may typically pass through levels of consciousness: from coma and vegetative states to a minimally conscious state. The coma state is characterized by the absence of sleep-wake cycles, no voluntary movement or behavioral response, and no response to auditory or visual stimuli.¹² In general, this progresses to a vegetative state with a return of sleep-wake cycles. In both the vegetative and coma states, there is no interaction with the environment or evidence of awareness of the environment. Positron emission tomography activation studies suggest that the vegetative state represents a global disconnection syndrome in which

higher order association cortices are functionally disconnected from primary cortical areas. In contrast, patterns of activation in functional magnetic resonance imaging studies of patients in the minimally conscious state show preservation of large-scale cortical networks associated with language and visual processing.¹³ Medical and rehabilitative interventions here focus on maintaining organ integrity and awaiting improvements in neurological status.

Careful examination is often needed to distinguish coma and vegetative states from minimally conscious states. The minimally conscious state is defined as "a condition of severely altered consciousness in which minimal but definite behavioral evidence of self or environmental awareness is demonstrated."¹² Physical examination criteria to establish awareness and interaction with the environment have been established.¹⁴ This level of baseline tonic arousal, which is defined as the ability to respond to the environment, is necessary in order for a patient to achieve maximal functional recovery from a comprehensive rehabilitation program; at this point, voluntary interaction on the part of the patient with his therapeutic team can begin.

Pharmacological neurostimulation may be used to increase a TBI patient's baseline tonic arousal level.¹⁵ If a patient is exhibiting signs of hypoarousal, it may be due to his TBI, but other factors must be ruled out before a pharmacological neurostimulation program is initiated. Factors such as mental/physical fatigue, post-TBI sleep/wake cycle inversion, premorbid personality issues (eg, stubbornness), and other psychological factors (eg, depression) can present a hypoarousal-like picture. Behavioral strategies for arousal problems include appropriate pacing of therapies as well as frequent breaks and naps.

A patient's ability to focus his attention on the therapeutic tasks at hand is also necessary to obtain results. Attention is defined as the ability to focus on a specific stimulus while other stimuli are ignored.¹⁶ A number of behavioral strategies can be used. These include the creation of a low-stimulation environment to prevent distractions from, for example, noises, hospital activities, and other people. Neurostimulants can also be used to improve attention. In a randomized controlled study of 34 patients, treatment twice daily with methylphenidate seemed to have clinically significant positive effects on speed of processing, caregiver ratings of attention, and some aspects of on-task behavior in naturalistic tasks.¹⁷ Another study of 16 inpatients with TBI and subarachnoid hemorrhage revealed significant improvement in attention and divided attention.¹⁸ Ongoing research continues to search for additional

ways to stimulate a patient's attention and to confirm the efficacy of current techniques.

NEUROREHABILITATION: IMPACT ON OUTCOMES

Although neurorehabilitation programs for patients with TBI are used, to date there are no definitive protocols for treatment of the multiple aspects of brain injury. Because providing neurorehabilitation after a TBI has become standard, ethical issues arise in finding or creating a control group that is not given neurorehabilitation after TBI. As described previously, neurorehabilitation is a tailored treatment. Currently the timing, intensity and components that optimize outcomes after TBI are not well understood.

The uniqueness of each patient's constellation of injuries and symptoms and, importantly, how they manifest with respect to his return to his previous life can make it difficult to determine the effects of therapeutic interventions. Goals in brain injury rehabilitation include issues such as returning to parenting and a previous career. These types of outcomes are dependent on the intimate interrelation between physical, cognitive, and behavioral/emotional issues, which are difficult to assess independently. For example, physical function will progress better in a patient with less cognitive and emotional impairment because he will be able to engage more in his physical therapy program.

In 2005, the *Cochrane Review* published an article entitled "Multi-Disciplinary Rehabilitation for Acquired Brain Injury in Adults of Working Age."¹⁹ Turner-Stokes *et al.*¹⁹ searched the medical literature for controlled trials comparing 2 groups of patients who received different treatments. They identified "ten trials of good methodological quality and four of lower quality." In their evaluation, the authors made a distinction between mild and moderate-to-severe injury and the evidence supporting the efficacy of multidisciplinary treatment. They were able to state that within the mild brain injury group, "strong evidence suggested that most patients made a good recovery with provision of appropriate information, without additional intervention." This is consistent with data showing that 85% of patients after a TBI of mild severity recover. Treatments for this group may be educational and supportive to manage their symptomatology. Fifteen percent of patients, sometimes called the miserable minority, will need a tailored rehabilitation program to maximize their function.

In patients who had sustained a moderate-to-severe injury, they found "strong evidence of benefit from formal intervention."¹⁹ They also reported that for patients with moderate-to-severe brain injury already in rehabilitation, there was strong evidence that more intensive programs were associated with earlier functional gains and moderate evidence that "continued outpatient therapy can help to sustain gains made in early post-acute rehab."¹⁹ An important point that they made is that "rehabilitation for brain injury is such an individual and long-term process that it can be difficult to draw general conclusions from research studies."¹⁹

A systematic review of the rehabilitation literature of moderate-to-severe acquired brain injuries was performed by Teasell *et al.* in 2007.²⁰ They found 275 interventional studies, but only 76 (28%) were randomized controlled clinical trials. They concluded that there was "need for studies of improved methodological quality into acquired brain injury rehabilitation."²⁰ A study to evaluate motor impairment rehabilitation after acquired brain injury was performed by Marshall *et al.* and published in 2007.²¹ They investigated pharmacological, nonpharmacological, and exercise interventions available for post-TBI motor impairments. They concluded that "although there are a variety of treatment strategies to manage motor impairments and activity limitations following acquired brain injury, most are only supported by limited evidence pointing to the need for studies of improved methodological quality in this area."²¹ The aforementioned reviews all conclude that additional research for verification of current treatments is necessary not only to confirm the effectiveness of those therapies but also to trigger and inspire investigation into new therapies for TBI patients. The heterogeneity of the injury and population and the multitude of ways to measure functional outcome in the TBI population make it uniquely challenging to determine the effectiveness of therapeutic interventions.

An exciting new research method is being used to study the efficacy of inpatient rehabilitation interventions after TBI. Practice-based evidence is a research method designed to observe naturally occurring differences in standard practice and identify particular patients for whom specific interventions are effective.²² Practice-based evidence/clinical practice improvement harnesses the complexity of patient and treatment differences in the actual practice of care. Unlike a randomized controlled trial, it does not alter the treatment to evaluate the efficacy of a particular intervention. Instead, it captures in-depth, comprehensive information about patient characteristics, processes of care, and outcomes to

Table 5. *Features of Practice-Based Evidence Studies.*

All interventions are considered to determine the relative contribution of each.
 Hypotheses are general.
 Minimal patient selection criteria maximize generalizability and external validity.
 A detailed characterization of patients is made through the use of robust measures of severity of illness and functional status.
 Patient differences are controlled statistically rather than through randomization.

characterize the process of care and ascertain the contribution of individual processes to outcomes, controlling for patient differences.²² The features of practice-based evidence studies are presented in Table 5.

The National Institutes of Health are funding a study of rehabilitation after TBI. This will include 12 neurorehabilitation centers in the United States and Canada, enrolling over 2000 patients over a 2.5-year period. Data will be collected from physicians, nurses, physical therapists, occupational therapists, recreational therapists, speech-language pathologists, social workers, and psychologists. Data will be collected on an inpatient basis and then at 3 months, 6 months, and 1 year post-injury. The study is designed to determine the extent to which

- Differences in severity of injury explain variations in outcomes following acute rehabilitation for TBI.
- Differences in patient demographic characteristics and severity of illness (both complications and comorbidities) explain variations in outcome from acute rehabilitation.
- Specific medical procedures and therapy interventions are associated with better outcomes.
- Levels of impairment interact with treatment activities to produce better outcomes.

CONCLUSION

Although most mild TBI patients do improve spontaneously, moderately and severely brain-injured patients are likely to benefit from a comprehensive rehabilitation program (Table 6). Because of the heterogeneity of injury, it will continue to be difficult to develop a standard TBI treatment protocol. There is still a great amount of work to be done with respect to the development of new pharmacological and therapeutic treatments. Also, the traditional approaches to TBI continue to be evaluated for their efficacy.

Table 6. *Key Concepts in Neurorehabilitation.*

The devastating effects of moderate and severe traumatic brain injury demand a multidisciplinary treatment program to maximize functional recovery.
 Early referral to a physiatrist to initiate a comprehensive rehabilitation program will result in the best outcomes.
 Although many traditional rehabilitative approaches have not yet been proven, they do improve functional recovery.
 Recovery from a brain injury does not proceed in a smooth progressive return of function. It is rather found to be a series of functional improvements interspersed with periods of stagnancy and worsening symptoms.
 Additional research is warranted to establish rehabilitative techniques that promote maximum functional recovery.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Langlois JA, Rutland-Brown W, Wald M. The epidemiology and impact of traumatic brain injury: a brief overview. *J Head Trauma Rehabil* 2006; 21: 375–378.
2. Bruns J, Hauser WA. The epidemiology of traumatic brain injury: a review. *Epilepsia* 2003; 44(suppl 10): 2–10.
3. Hibbard MR, Bogdany J, Uysal S, et al. Axis II psychopathology in individuals with traumatic brain injury. *Brain Inj* 2000; 14: 45–61.
4. Ponsford JL, Olver JH, Curran C. A profile of outcome: 2 years after traumatic brain injury. *Brain Inj* 1995; 9: 1–10.
5. Annegers JF, Hauser WA, Coan SP, Rocca WA. A population-based study of seizures after traumatic brain injuries. *N Engl J Med* 1998; 338: 20–24.
6. Bratton SL, Chestnut RM, Ghajar J, et al. Guidelines for the management of severe traumatic brain injury. XIII. Antiseizure prophylaxis. *J Neurotrauma* 2007; 24(suppl 1): S83–S86.
7. Englander J, Bushnik T, Duong TT, et al. Analyzing risk factors for late posttraumatic seizures: a prospective, multicenter investigation. *Arch Phys Med Rehabil* 2003; 84: 365–373.
8. Watanabe TK, Sant MO. Common medical complications of traumatic brain injury. *Phys Med Rehabil* 2001; 15: 283–299.
9. Cifu DX, Kaelin DL, Wall BE. Deep venous thrombosis: incidence on admission to a brain injury rehabilitation program. *Arch Phys Med Rehabil* 1996; 77: 1182–1185.
10. Geerts WH, Code KI, Jay RM, et al. A prospective study of venous thromboembolism after major trauma. *N Engl J Med* 1994; 331: 1601–1606.
11. Ghigo E, Masel B, Aimaretti G, et al. Consensus guidelines on screening for hypopituitarism following traumatic brain injury. *Brain Inj* 2005; 19: 711–724.

12. Giacino JT, Ashwal S, Childs N, et al. The minimally conscious state, definition and diagnostic criteria. *Neurology* 2002; 58: 349–353.
13. Giacino JT, Hirsch J, Schiff N, Laureys S. Functional neuroimaging applications for assessment and rehabilitation planning in patients with disorders of consciousness. *Arch Phys Med Rehabil* 2006; 87(suppl 2): S67–S76.
14. Giacino JT. The vegetative and minimally conscious states: consensus-based criteria for establishing diagnosis and prognosis. *NeuroRehabilitation* 2004; 19: 293–298.
15. Napolitano E, Elovic EP, Qureshi AI. Pharmacological stimulant treatment of neurocognitive and functional deficits after traumatic and non-traumatic brain injury. *Med Sci Monit* 2005; 11: 212–220.
16. Whyte J. Attention and arousal: basic science aspects. *Arch Phys Med Rehabil* 1992; 73: 940–949.
17. Al-Adawi S, Calvanio R, Dorvlo A, et al. The effect of methylphenidate on attention in acquired brain injury as recorded by useful field of view. *J Appl Res* 2005; 5: 2005.
18. Whyte J, Hart T, Vaccaro M, et al. Effects of methylphenidate on attention deficits after traumatic brain injury: a multidimensional, randomized, controlled trial. *Am J Phys Med Rehabil* 2004; 83: 401–420.
19. Turner-Stokes L, Disler PB, Nair A, Wade DT. Multidisciplinary rehabilitation for acquired brain injury in adults of working age. *Cochrane Database Syst Rev* 2005; 3: CD004170.
20. Teasell R, Bayona N, Marshall S, et al. A systematic review of the rehabilitation of moderate to severe acquired brain injuries. *Brain Inj* 2007; 21: 107–112.
21. Marshall S, Teasell R, Bayona N, et al. Motor impairment rehabilitation post acquired brain injury. *Brain Inj* 2007; 21: 133–160.
22. Horn SD, Gassaway J. Practice-based evidence study design for comparative effectiveness research. *Med Care* 2007; 45: S50–S57.

Essay: The Future of Traumatic Brain Injury

Jamshid Ghajar, MD, PhD

Weill Cornell Center and the Brain Trauma Foundation, New York, NY

The worldwide incidence of traumatic brain injury (TBI) is steadily increasing as a result of increasing road accidents and better diagnosis of mild TBI. The increase in brain injuries comes with an enormous cost to individuals, their families, and society. The good news is that advances have been made, especially in our treatment of severe TBI, and imaging in mild TBI. Future advances in TBI will depend on regional trauma centers that cover the TBI spectrum from prevention to rehabilitation and multi-institutional clinical data collection for scientific analysis to improve the diagnosis and treatment of TBI.

The real cure for the TBI epidemic is prevention, and this must be the primary focus for public health initiatives. Advances have been made in both helmets and restraint systems, which have decreased the incidence and severity of injury. However, current helmet designs and restraint systems are focused on decreasing injury from impact rather than preventing brain shearing from angular acceleration/deceleration; this shearing results from rapid head and neck movement and contributes significantly to the cognitive deficits seen in TBI.¹ Large, bulky helmets may actually accentuate this angular, rotational shearing. The incidence of brain injury in the future could be dramatically decreased with a combination helmet/torso restraint system that prevents both focal and diffuse brain injury. Public health strategies that promote the use of these systems is of fundamental importance.

When a brain injury does happen the medical triage is based on the level of arousal with coma, severe TBI, as the most urgent condition because of

possible systemic and intracranial treatable pathologies. Concussion or mild TBI as a result has been neglected in diagnosis and treatment because of the significant difference in outcome compared to the TBI patient in coma, the difficulty in diagnosis and because the majority of patients make a functional recovery. However the incidence of mild TBI is a magnitude greater than severe TBI and because of subtle attention deficits, these post-concussive individuals are at risk for recurring injuries. Concussion and coma, while sharing a continuum of diffuse axonal injury, currently have different diagnostic and treatment issues that need to be addressed separately.

CONCUSSION

The focus of public attention and medical specialists has shifted from severe TBI to the proportionately much larger group of patients classified as having concussion (mild TBI). The wars in Iraq and Afghanistan have generated an overwhelming number of these cases, which have heightened awareness and concern that mild TBI might not be so mild; indeed, there is growing evidence that mild TBI may be associated with both short-term and long-term sequelae that must be recognized and addressed. Sports, assaults, and motor vehicle accidents contribute to the growing number of patients sustaining mild TBI in the United States; the number of cases is increasingly at an even faster rate in developing countries as a result of the availability of motorized vehicles. Although there is little controversy about the diagnosis of coma in severe TBI, concussion diagnosis remains mired in various expert opinions without a firm scientific basis.

The core cognitive deficit in concussion is attention, which leads to working memory dysfunction and subsequent secondary symptoms and compensation. Without a scientific understanding of how we pay attention, there can be no consistent, reliable diagnosis or therapy for concussion.

Address Correspondence to:

Jamshid Ghajar, MD, PhD
Weill Cornell Medical College
New York, NY
Email: ghajar@braintrauma.org

Therefore, a fundamental question for constructive mild TBI research is whether attention can be scientifically defined and measured. One hypothesis with significant scientific backup is that attention can be defined as a predictive brain state.² Sustained attention is a luxury of a predictable environment; that is, we know what to expect and match or synchronize that with sensory input. Although we are aware and focus on the "what," the "when" occurs unconsciously: content is explicit, but timing is implicit. I listen to the word content (explicit) in a cell phone conversation but not to the timing of the vocalizations (implicit).

Disruption of this predictive brain state is postulated as a cause of symptoms in concussion resulting from white matter shearing in anterior and lateral fibers that connect the prefrontal cortex with parietal and cerebellar regions—regions that subserve the attention neural network.³ Recent imaging studies using magnetic resonance diffusion tensor imaging (DTI) in concussion patients have revealed white matter integrity changes in the anterior lateral frontal region, which is the region most vulnerable to angular acceleration/deceleration forces.⁴ The degree of white matter integrity change after a concussion correlates well with attention and working memory measures according to whether the lesion is anterior frontal or lateral frontal, respectively.⁵

A diagnosis of attention dysfunction following a concussion should be based on a cognitive timing function according to the premise of the predictive brain state. Traditionally, reaction times or other processing speed parameters are used as part of cognitive assessment, with the variability of the timing parameter being a hallmark of TBI.⁶ However, reaction time testing is intermittent and slow, missing lapses in attention. A faster and more continuous assessment is tracking the eye following a predictable target, as in the circular smooth pursuit task.⁷ Synchronization or variability measures can be obtained in seconds and correlate with attention and working memory measures and DTI lesions in concussion patients. The Department of Defense recently funded the further development of this timing assessment for military use.⁸ Therapeutic potential for improving attention may come from direct feedback of timing variability, which could make an implicit deficit explicit and malleable.

The diagnosis of a mild TBI with improved imaging techniques such as DTI is a moving target that is based on a comparison with normative data. Analysis of DTI involves thresholds for injury, methods of white matter integrity, and controls for aging effects, all factors that confound the diagnosis.

Also, the age of DTI lesions and changes over time are still unknowns.

The term *functional brain injury* is needed to define those that have white matter abnormalities referable to a functional deficit because many normally functioning individuals will be found to have lesions as imaging techniques improve. Also, we will find that a significant number of people are "the walking wounded," that is, individuals who through previous TBI or degeneration have DTI lesions in the attention neural network with concomitant attention deficits and are prone to injuries, including recurrent TBI. The bottom line is that as imaging techniques improve, the definition of *normal* needs to be ascertained with functional measures and not imaging data alone.

COMA

In the last 20 years, mortality from severe TBI has fallen from approximately 55% to 20%. What is most impressive in this improvement in outcome is that it has occurred in the absence of any new pharmaceutical or other novel therapeutic interventions.^{9,10} This remarkable progress has been largely achieved by maintenance of brain perfusion from the accident scene through the intensive care unit course by full resuscitation of patients and reduction of intracranial pressure without brain perfusion being sacrificed.^{11,12}

Current recommendations emphasize simple maneuvers to equilibrate blood pressure and brain pressure. These maneuvers are exactly the opposite of the treatment protocol of 20 years ago: "pruning" out the patient by reducing fluid volume with the liberal use of mannitol, fluid restriction, and the constriction of brain perfusion with prophylactic hyperventilation was once the standard of care established by consensus but has since been refuted with evidence. The resulting hypotension was, and continues to be, a major cause of mortality and morbidity in severe TBI.

During these past 20 years, over 25 therapeutic trials of new pharmaceuticals, costing billions of dollars to develop, have failed to prove efficacy in randomized clinical trials.¹³ Many of these novel interventions are based on our understanding of injury mechanisms and are theoretically sound with supporting evidence from animal models. Indeed, it is unclear why many of the trials have failed, although contributing factors may include the timing of the intervention and clinical practice variability.

TRAUMATIC BRAIN INJURY RESEARCH

Although basic research on mechanisms of TBI pathology and novel therapeutic developments add to our scientific knowledge, it seems that the major advances in improving outcomes have come from observational clinical science rather than laboratory science: This finding should cause reflection and brings into question the current research approach. The evidence suggests that future improvements in TBI outcomes will be a result of questioning and changing our current structure of TBI research and clinical care. Perhaps it is time to rethink how TBI research is being conducted and to consider shifting our efforts in the direction of observational clinical research. This shift would require accurately collected, relevant parameters and subsequent analysis for evidence-based guidelines. Also, pooling of patients in regionalized trauma centers would provide uniform, standardized care and the optimal setting for observational science, allowing for single-center clinical trials and rapid, high-volume referrals to specialized TBI rehabilitation centers. These practical steps leading to improvements in outcomes are embodied in the scientific evidence-based guidelines loop (Figure 1).

THE FUTURE

There are many challenges yet to be met in preventing, diagnosing, and treating TBI. Prevention must be the primary focus and can be facilitated by both public education initiatives (eg, fall prevention strategies for the elderly) and applications of technologies (eg, advanced helmet-torso design). Early and full resuscitation has had the greatest impact on improving outcome from severe TBI; indeed in many ways the

future is already here by ensuring oxygenation and perfusion of the injured brain.

Research on new therapeutics must shift toward greater emphasis on clinical data collection. The recent creation of neurological research networks such as the Neurologic Emergencies Treatment Trials Network (sponsored by the National Institute of Neurologic Diseases and Stroke) will provide the infrastructure needed for multicenter research that incorporates both prehospital systems and trauma systems.¹⁴

Perhaps the most exciting frontier in TBI research is gaining a better understanding of the neurobehavioral sequelae of TBI. New imaging technologies combined with neuropsychological and neurophysiological testing hold the potential to help identify lesions that contribute to neurobehavioral deficits, which often complicate recovery. Focusing research on attention as the signature cognitive deficit in TBI should lead to improved diagnostic and treatment strategies. The scientific community of TBI researchers and clinicians is now lifted on a wave of public awareness and support as a result of the war and reporting on TBI from sports injuries. We have the tools to successfully meet these challenges and significantly improve the lives of TBI patients.

DISCLOSURES

Potential conflict of interest: U.S. Patent Number 7,384,399, Cognition and Motor Timing Diagnosis and Training System and Method.

REFERENCES

1. Arash A, Sabet EC, Benjamin Z, et al. Deformation of the human brain induced by mild angular head acceleration. *J Biomech* 2008; 41: 307–315.
2. Ghajar J, Ivry R. The predictive brain state: asynchrony in disorders of attention? *Neuroscientist* 2008; doi://10.1177/1073858408326429. Available at: <http://nro.sagepub.com>.
3. Ghajar J, Ivry R. The predictive brain state: timing deficiency in traumatic brain injury? *Neurorehabil Neural Repair* 2008; 22: 217–227.
4. Niogi S, Mukherjee P, Ghajar J, et al. Extent of microstructural white matter injury in postconcussive syndrome correlates with impaired cognitive reaction time: a 3T diffusion tensor imaging study of mild traumatic brain injury. *ANJR Am J Neuroradiol* 2008; 29: 967–973.
5. Sumit N, Mukherjee P, Ghajar J, et al. Structural dissociation of attentional control and memory in adults with and without mild traumatic brain injury. *Brain* 2008; 131: 3209–3221.
6. Stuss D, Pogue J, Buckle L, Bondar J. Characterization of stability of performance in patients with traumatic

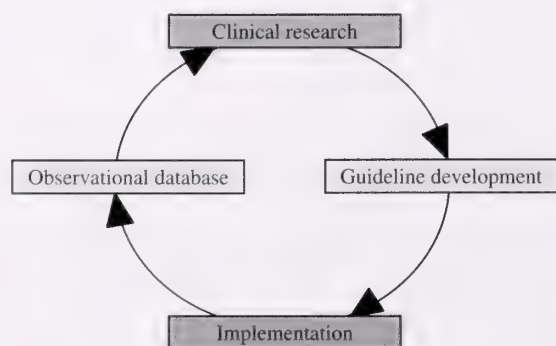


Fig 1. Scientific evidence-based guidelines loop. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

- brain injury: variability and consistency on reaction time test. *Neuropsychiatry* 1994; 8: 316–324.
7. Suh M, Kolster R, Sarkar R, et al. The cognitive and neurobiological research consortium: deficits in predictive smooth pursuit after mild traumatic brain injury. *Neurosci Lett* 2006; 401: 108–113.
 8. Francis T. Pentagon Seeks Battlefield Device to Diagnose Brain Injury. *Wall Street Journal Health Blog*, April 17, 2008.
 9. Ghajar J. Traumatic brain injury. *Lancet* 2000; 356: 923–929.
 10. Hesdorffer DC, Ghajar JBG. Marked improvement in adherence to traumatic brain injury guidelines in United States trauma centers. *J Trauma* 2007; 63: 841–848.
 11. Bratton S, Chesnut RM, Ghajar JBG, et al. Guidelines for the management of severe traumatic brain injury, 3rd edition. *J Neurotrauma* 2007; 24(suppl 1): 1–200.
 12. Faul M, Wald M, Rutland-Brown W, et al. Using a cost-benefit analysis to estimate outcomes of a clinical treatment guideline: testing the Brain Trauma Foundation guidelines for the treatment of severe traumatic brain injury. *J Trauma* 2007; 63: 1271–1278.
 13. Narayan RK, Michel ME, Ansell B, et al. Clinical trials in head injury. *J Neurotrauma* 2002; 5: 503–557.
 14. Cearnal L. The birth of the NETT: NIH-funded network will launch emergency neurological trials. *Ann Emerg Med* 2006; 48: 726–728.

EVIDENCE-BASED MEDICINE

Clopidogrel for Preventing Cardiovascular Events

James Stulman, MD, Thomas McGinn, MPH, MD and Deborah Korenstein, MD

Mount Sinai School of Medicine, New York, NY

This is the fourth in a series of articles in the *Mount Sinai Journal of Medicine* on evidence-based medicine; this series will emphasize applications of evidence-based medicine concepts and procedures to clinical practice. The articles will focus on the retrieval, interpretation, and application of evidence related to a current clinical problem. The discussion will be based on the following case.

THE CASE

Your patient is a 64-year-old female with a history of hypertension and type 2 diabetes with nephropathy who has smoked 1 pack of cigarettes daily for over 30 years. She comes to you for a routine follow-up visit. Her blood pressure and diabetes are well controlled. Despite numerous efforts, the patient has not been able to quit smoking. Her medications include an angiotensin-converting enzyme inhibitor, a diuretic, a statin, metformin, and aspirin. You know that she is at increased risk for a myocardial infarction and stroke. You wonder if dual antiplatelet therapy with a combination of clopidogrel and aspirin will offer the patient greater protection from cardiovascular events.

FOCUSED CLINICAL QUESTION

To determine if adding clopidogrel to aspirin improves outcomes, you decide to review the

relevant medical literature. You begin by formulating a focused question consisting of 4 elements: the patient (P), the intervention (I), the comparison (C), and the outcome (O). This is called a PICO question.¹ The explicit inclusion of these 4 components allows for a more efficient and directed search of the literature and helps to clarify the exact problem of interest. In this case, the question's components are as follows: in patients at high risk for atherothrombosis (P), is treatment with clopidogrel plus aspirin (I), in comparison with aspirin alone (C), more effective in reducing myocardial infarction and stroke (O).

SEARCH OF THE LITERATURE

You search Medline using the PubMed interface by going to the PubMed Web site (<http://www.ncbi.nlm.nih.gov/pubmed>). In the left-hand column, you select "Clinical Queries." This filter is designed specifically for clinicians. This search tool incorporates 2 specific search filters that will assist you in finding the best available evidence for your clinical questions: (1) you can focus the electronic search of the literature by choosing a specific clinical study category (eg, therapy, diagnosis, or prognosis), and (2) you can determine the scope of your search by choosing either a narrow, specific search or a broad, sensitive search. The Clinical Queries search engine automatically adds terms to each search to find articles with the best study design to answer questions in each category. For example, in the therapy category, Clinical Queries searches will find only randomized controlled trials. When you are asking a clinical question that has been addressed by high-quality studies, Clinical Queries will allow you to quickly access the best data.

In the search box, you enter the following keywords, which are contained in your well-formulated clinical question: clopidogrel, aspirin, myocardial infarction, and stroke. Next, you use the

Address Correspondence to:

James Stulman, MD
Mount Sinai School of Medicine
New York, NY
Email: james.stulman@mountsinai.org

aforementioned specially designed search filters. As this is a question regarding therapeutics, you select "therapy" in the category column. You expect to find a large number of publications on this general topic and therefore select "narrow" in the "scope" column. The result of this initial search is 53 citations. To focus your search on the most relevant articles, you click the "limits" button and select English. The result of this search is 44 articles. To focus the search even more, you select "limits" again and select articles written in the past 5 years and core journals. The result of this new search is 22 citations. You review this manageable list and select #10, "Clopidogrel and Aspirin Versus Aspirin Alone for the Prevention of Atherothrombotic Events," by Bhatt and colleagues, which was published in the *New England Journal of Medicine* in April 2006.² You review this article.

CRITICAL APPRAISAL OF THE LITERATURE

Critical appraisal has 3 main components: an assessment of the validity of the trial, an analysis of the results, and a review of the applicability.

The Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management, and Avoidance (CHARISMA) trial was a pharmaceutical company-sponsored trial. The study recruited patients from 32 countries and 768 sites. The patients had to be 45 years of age or older and have 1 of the following conditions: multiple atherothrombotic risk factors, documented coronary artery disease, documented cerebrovascular disease, or documented symptomatic peripheral vascular disease. Patients were excluded if they were taking oral antithrombotic medications or nonsteroidal anti-inflammatory drugs on a long-term basis. Patients were also excluded if they had an established indication for clopidogrel.

The primary endpoint for the trial was the first occurrence of myocardial infarction, stroke of any cause, or death from cardiovascular causes. The principle secondary outcome was the above plus hospitalization for unstable angina, transient ischemic attack, or a revascularization procedure. There were other endpoints, including nonfatal stroke and death from any cause. The primary safety endpoints were severe bleeding and moderate bleeding.

Validity Criteria for Treatment Trials

Before looking at the results of the study, you assess the validity of the study design. When examining the validity of the trial, you are asking if the methodology will produce results with a minimum of bias. If at the

end of this section you conclude that there is a significant amount of bias in the methods, then the results will not be believable.

The study was randomized. The patients were randomly assigned to receive either clopidogrel and aspirin or aspirin and a placebo.

Allocation concealment was performed adequately. The authors report that "study drug assignment was performed centrally by an interactive voice-response system on a basis of a pre-established randomization scheme. . . ." Thus, it is unlikely that physicians enrolling a patient could predict to which group the patient would be assigned.

The patients in the 2 groups were similar with respect to known prognostic factors. The 2 groups of patients were statistically similar in terms of demographics, clinical characteristics, and medications taken.

The groups were treated equally. All patients received standard therapy as appropriate and received the same follow-up evaluations.

Patients who entered the trial were accounted for. Data regarding the primary endpoint were known in 99.5% of the patients randomized to clopidogrel and aspirin and in 99.6% of the patients randomized to aspirin and placebo. Therefore, the number of patients lost to follow-up was very small.

The study was performed and analyzed in an intention-to-treat (ITT) fashion. This means that all patients were analyzed in the groups to which they were originally assigned. ITT analysis preserves the most important feature of this study design, namely, the randomization of the patients. However, when a large number of patients cross over to the other group, the ITT analysis may obscure significant differences between the 2 groups. Similarly, large discontinuation rates can also alter interpretation of the ITT analysis. Sometimes, a secondary, as-treated analysis is performed to see if there are significant differences with the ITT analysis. In addition, crossover rates were not reported in this study; however, it is unlikely, given the nature of this study, that patients did cross over to other treatment groups. The authors did report that assigned treatment was permanently discontinued by 20.4% of the patients in the clopidogrel group versus 18.2% of those patients receiving the placebo ($P < 0.001$). Although this was statistically significant, the authors did not provide an as-treated analysis to measure the effect of this sizeable discontinuation rate. The next question that the reader must ask is why patients discontinued their medications. More specifically, did adverse events lead to a disparity in dropout rates in one group versus the other? In this trial, 4.8% of the patients in the clopidogrel group and 4.9% of the patients

receiving the placebo discontinued treatment because of adverse events ($P = 0.67$). It appears that adverse events differed between the groups.

Both patients and clinicians were blinded to the treatment being received. In addition, data collectors, data analysts, and safety monitors were blinded.

Results

You have now completed a systematic review of the methodology, which reveals no major concerns regarding the introduction of bias. You therefore conclude that the design of the study is valid and thus will produce believable results. You now analyze the results of the study in more detail.

The CHARISMA trial assigned 7802 to the clopidogrel plus aspirin group and 7801 to the aspirin plus placebo group. Median follow-up was 28 months. For the primary composite endpoint (myocardial infarction, stroke, or death from cardiovascular causes), there was no difference between the groups. The rate was 6.8% for the clopidogrel group and 7.3% for the placebo group [relative risk (RR) = 0.93, 95% confidence interval (CI) = 0.83 – 1.05, $P = 0.22$].

When the data are provided, it is always important to look at the rates of “death from any cause.” This endpoint is the most objective and carries a significant value to patients and physicians. For this study, the rate of death from any cause was 4.8% in both groups.

For the secondary composite endpoint (the primary composite plus hospitalization for unstable angina, transient ischemic attack, or a revascularization procedure), there was a slight risk reduction for the clopidogrel group: 16.7% versus 17.9% (RR = 0.92, CI = 0.86–0.995, $P = 0.04$). Another way to appreciate and interpret risk reduction is to calculate the number needed to treat (NNT). In order to calculate the NNT, we must first calculate the absolute risk reduction (ARR):

$$\text{ARR} = \text{Incidence outcome in the control group} \\ - \text{Incidence outcome in the study group}$$

For this study, the ARR for the secondary composite outcome was 1.2% (ie, 17.9% – 16.7%). The ARR most clearly demonstrates the effect of the intervention on a population. The NNT tells you how many patients need to be treated with a specific intervention in order to prevent 1 outcome. You can calculate the NNT by taking the reciprocal of the ARR or $1/\text{ARR}$. For this study, the NNT was 83 (ie, $1/0.012$). In other words, 83 patients in this trial needed to be treated with clopidogrel in order to prevent 1 of the secondary outcomes.

Similarly, for the endpoint of nonfatal stroke, there was a slight benefit seen in the clopidogrel group (RR = 0.79, CI = 0.64–0.98, $P = 0.02$). The NNT to prevent 1 stroke was determined by the calculation of $1/\text{ARR}$ or $1/(2.4\% - 1.9\%) = 200$. Once more, this means that in order to prevent 1 nonfatal stroke, 200 patients would be needed to be treated with clopidogrel.

Looking at the safety endpoints, you find no significant differences between the 2 groups in the rate of severe bleeding episodes, but there was an increase in moderate bleeding episodes in the clopidogrel group: 2.1% versus 1.3% (RR = 1.62, CI = 1.27–2.08, $P > 0.001$). To interpret these data in a clinically useful manner, we can calculate the number needed to harm (NNH). The NNH is similar to the NNT but is used to express adverse events rather than beneficial outcomes. The NNH tells you how many patients in a study need to be exposed to the intervention in order to produce 1 single adverse outcome. The first step in calculating the NNH is to calculate the absolute risk increase (ARI). This is done by subtraction of the rates of moderate bleeding observed in each group: $2.1\% - 1.3\% = 0.08\%$ (or 0.008). The NNH is then calculated by $1/\text{ARI}$ or $1/0.008 = 125$. Thus, in this study, for every 125 patients who received clopidogrel and aspirin, 1 patient suffered moderate bleeding.

APPLICABILITY

You determine that this published study, the CHARISMA trial, does apply to your patient because she meets the study's entry criteria. You decide that despite the patient's increased risks for atherothrombotic events, the evidence does not support adding clopidogrel to her regimen in order to prevent myocardial infarction or stroke. You consider other outcomes, including the reduction in nonfatal stroke and the secondary composite endpoint, as a rationale for prescribing clopidogrel. However, you discuss with your patient the NNT of each endpoint along with the NNH for bleeding. You and the patient jointly conclude that the risks of bleeding offset the small benefits demonstrated in this trial.

DISCUSSION POINTS

1. Your knowledge of the pathophysiology of atherothrombotic disease was the starting point for generating a question regarding optimal care for your patient. Yet to safely treat patients, we typically

require a higher level of evidence than knowledge of pathophysiology. When available, a well-designed randomized controlled trial such as the CHARISMA trial offers powerful evidence to help guide our clinical decision making.

2. The Clinical Queries function on PubMed is an easy-to-use search tool specially designed for assistance in answering clinical question.
3. The use of composite outcomes can be confusing to interpret. Always make sure that, in studies measuring mortality, the authors report all-cause mortality. These data may uncover unforeseen effects of the intervention and are the most objective and therefore easiest to interpret.
4. When results are being interpreted, the ARR and NNT can be easily calculated. These results help convey the true clinical impact of a particular intervention. Similarly, the NNH expresses the risks posed by an intervention. Incorporating the

concepts of NNT and NNH will help you in your clinical decision making. Moreover, discussing the specific risks and benefits of an intervention will assist you and your patients in making more informed decisions together.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Friedland DJ, Go AS, Davoren JB, et al. *Evidence Based Medicine: A Framework for Clinical Practice*. New York, NY: Appleton & Lange; 1998.
2. Bhatt DL, Fox KA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N Engl J Med* 2006; 354: 1706–1717.

NEW DRUGS AND THE FDA

Cardiovascular Safety of Medication Treatments for Attention-Deficit/Hyperactivity Disorder

Jeffrey H. Newcorn, MD¹ and Craig Donnelly, MD²

¹Mount Sinai School of Medicine, New York, NY

²Dartmouth-Hitchcock Medical Center, Lebanon, NH

On April 21, 2008, a news release advised that a review conducted by a subcommittee of the American Heart Association (AHA) concluded that the guidelines for pretreatment medical work-up of children and adolescents with attention-deficit/hyperactivity disorder (ADHD) should be revised. The new guidelines, which were published in the journal *Circulation* the following day,¹ recommended that all children and adolescents should have a medical work-up that includes a baseline electrocardiogram (ECG) prior to beginning treatment with stimulant medication. Because stimulants are commonly prescribed treatments for ADHD that have been approved for over 45 years in one form or another and have generally been considered to have a favorable risk-benefit ratio, this announcement was indeed big news. It also signaled an apparent shift in the AHA guidelines, which had previously advised screening patients for preexisting conditions and a family cardiovascular history and monitoring pulse and blood pressure prior to and over the course of treatment but did not recommend that specific laboratory tests or ECGs be obtained.² Of note, the recommendation to obtain pretreatment ECGs would represent a departure from current labeling approved by the Food and Drug Administration (FDA) as well as guidelines issued by the American Academy of Pediatrics (AAP) and

the American Academy of Child and Adolescent Psychiatry (AACAP). In addition, the circumstances surrounding the manner in which this information was disseminated—via electronic and print media available to the general public and apparently without priority communication with the overwhelming majority of physician stakeholders and their professional organizations—seemed curious.

Did the AHA have new information that caused it to question the safety of existing treatments, or was this recommendation based on a reinterpretation of existing data? Moreover, what did the AHA consider the level of risk to be? After all, there are millions of children with ADHD who are currently prescribed psychostimulant medication, and the large majority almost certainly did not have ECGs either prior to or after initiation of treatment. Moreover, many pediatricians are not equipped to routinely obtain and/or interpret ECGs, and access to specialists in pediatric cardiology is often limited. Thus, the potential impact of this recommendation on clinical care was anticipated to be substantial. Would newly diagnosed patients decide to forego potentially effective treatment because of fear of cardiovascular risk (previously considered to be low)? Might patients currently being treated successfully with stimulants decide to discontinue medication precipitously?

Reaction to the news release and the *Circulation* article was swift. The AAP issued a statement asserting that the recommendation to obtain an ECG before starting stimulant medication for ADHD contradicted the carefully considered and evidence-based recommendations of the AAP and AACAP, which had previously concluded that sudden cardiac death in youth taking medications for ADHD is a very rare event and that there is no evidence that obtaining

Address Correspondence to:

Jeffrey H. Newcorn

Mount Sinai School of
Medicine, New York, NY

Email: jeffrey.newcorn@
exchange.mssm.edu

a screening ECG would appropriately balance issues related to benefit, risk, and cost effectiveness in the identification of risk factors for sudden death in children with ADHD who are treated with stimulants. The statement further indicated that, until these questions can be answered, "a recommendation to obtain routine ECGs for children receiving ADHD medications is not warranted."

BACKGROUND INFORMATION

Overview

ADHD is a highly prevalent neurodevelopmental disorder that begins early in life and often persists (in full or residual forms) across the lifespan (for a review, see Newcorn *et al.*³). Although it has often been dismissed as "boys being boys," ADHD is a condition with a high degree of heritability—roughly equivalent to that for schizophrenia and bipolar disorder and substantially higher than that for nonbipolar depression and several of the anxiety disorders. It is often highly impairing and carries adverse functional consequences related to academic achievement/occupational attainment, family/peer relationships, psychological development, and behavioral function. Findings from longitudinal and epidemiologic studies indicate that ADHD is associated with increased risk for a wide range of high-impact behaviors in a subgroup of individuals, including unsafe driving and sexual behavior, substance abuse, and antisocial disorders. Advances in the scientific basis of ADHD and its treatment have been bolstered by findings from pharmacological, neuroimaging, and animal studies, which implicate dopaminergic and noradrenergic neural circuits centered in the prefrontal cortex, striatum, cerebellum, and brainstem nuclei in the pathophysiology of the disorder. A plethora of approved medication treatments are available, including several formulations of methylphenidate and amphetamine (2 classes of psychostimulants) and atomoxetine (nonstimulant), all of which directly affect noradrenergic and/or dopaminergic neurotransmission. Psychosocial interventions are effective but are generally less effective in producing symptom reduction. For many patients, combined behavioral and pharmacological treatment represents the best option.

Prevalence, Prescription Rates, and Other Sources of Controversy

Recent epidemiologic data indicate that the prevalence of ADHD in children (approximately 7%–8%)

and adults (4%–5%) is considerably higher than what it was previously believed to be. The recent increase in rates of diagnosis and treatment in children has more or less stabilized. However, increased appreciation of the various ways in which ADHD can present in adults and improved methods of identification have led to corresponding increases in the rates of diagnosis and treatment and a corresponding increase in pharmacotherapy. These facts have been widely publicized in the scientific literature and popular press, in some cases drawing negative reactions. Yet, it is not clear to what extent these reactions might reflect attitudes toward the disorder itself (which is composed of symptoms that represent the extremes of behaviors that all people have), the use of medication to regulate behavior and enhance cognitive performance (which abuts controversial issues for our society), or the fact that psychostimulant medications are Schedule II controlled substances, which have the potential for abuse and diversion. An additional complexity is that there is no single laboratory test to determine whether a patient has ADHD and, if so, how best to treat it. Thus, although the neurobiological basis, longitudinal course, and disorder-related impairments of ADHD are extremely well documented and response to treatment is among the most robust in psychiatry, ADHD remains a somewhat controversial clinical syndrome.

Cardiovascular Safety of Attention-Deficit/Hyperactivity Disorder Treatments: State of Knowledge and Its Interpretation Prior to 2006

Because of their effects on noradrenergic receptors, it has long been known that stimulants (and more recently the nonstimulant atomoxetine) have adverse cardiovascular effects, the most frequent being mild increases in pulse and blood pressure. More pronounced changes in cardiovascular parameters can be seen early in treatment, for brief periods during the day, or in selected individuals. Arrhythmias or other more severe problems occur very infrequently, almost always in vulnerable patients with underlying cardiac abnormalities or in association with other medication treatments. In 1999, the AHA issued guidelines calling for the screening of patients for preexisting cardiac dysfunction or a family history of cardiac dysfunction and for the monitoring of pulse and blood pressure before and during treatment.² Screening should focus on the history of a cardiac event early in life or the early onset of known cardiovascular disease in either the subject

(eg, unexplained syncope, rhythm disturbance, or murmur) or a first-degree relative (eg, cardiac death prior to age 50 or a known structural cardiac abnormality). In addition, early, unanticipated cardiovascular risk and symptoms that emerge during exercise-related activities may indicate possible risk for cardiac dysfunction. Patients who screen positive should be referred to a pediatric cardiologist for more extensive examination using ECG, a Holter monitor, and/or echocardiogram, as indicated. The last may be particularly useful for evaluating possible structural outflow anomalies, which represent a specific risk factor for sudden death. The occurrence of periodic arrhythmias or changes in blood pressure should also be noted and evaluated appropriately.

RECENT EVENTS HIGHLIGHTING THE ISSUE OF CARDIOVASCULAR SAFETY OF STIMULANT TREATMENT

Temporary Withdrawal of Mixed Amphetamine Salts (Extended Release) by Health Canada

The issue of cardiovascular safety of stimulants came to public attention in 2006 when the popular medication Adderall XR (extended-release mixed amphetamine salts) was withdrawn from the Canadian market by Health Canada (the equivalent of the FDA).⁶ This decision was based on the occurrence of several sudden cardiac deaths identified in postmarketing surveillance in the United States. These data had been reviewed by the FDA in 2005,⁴ and a mid-level warning was added to the product insert. However, the magnitude of the problem was not seen as sufficiently high to withdraw the medication in the United States. The Health Canada decision was subsequently appealed, and after further review, the medication was reintroduced. The confusion and concerns that accompanied the Health Canada decision were most unfortunate; however, the additional level of scrutiny that it brought to this issue and the increased dissemination of important information regarding the cardiovascular safety of stimulants to prescribing physicians led to heightened awareness of the types of conditions that constitute high risk, more rigorous approaches to screening, and a higher level of consciousness regarding cardiovascular safety.

Food and Drug Administration Review of the Cardiovascular Safety of Stimulants

The FDA review of sudden cardiac death among patients taking stimulants for the treatment of ADHD followed the reporting of 25 cases of sudden cardiac death in 8 patients treated with methylphenidate (7 children and 1 adult) and 17 patients treated with amphetamine (12 children and 5 adults).^{5,6} Nonfatal cerebrovascular adverse events over the same 4-year period were reported in 8 youths on methylphenidate and in 18 on amphetamines.⁷ Yet, determining whether there is increased treatment-related risk can be evaluated only by an examination of the frequency of signal events against the base population rates and by sampling over a sufficiently large and representative time interval. From February 1992 to 2005, there were approximately 110 million prescriptions written for methylphenidate, covering a population of 7 million children; this resulted in an estimated rate of sudden cardiac death in the treated population that was 2 to 5 times below the corresponding rate in the general population.^{5,6} Thus, it did not appear that sudden death occurred more frequently in patients treated with stimulants. Moreover, a sizable number of the reviewed cases were determined to have had preexisting, though sometimes undetected, cardiovascular problems or risks or were using multiple drugs simultaneously.

On the basis of these and other data, the FDA Pediatric Psychopharmacology Advisory Panel concluded that there was no increased risk for sudden death in healthy patients with ADHD treated with stimulants.⁸ Consistent with the observation that untoward events mainly occur in the context of preexisting risk, the panel recommended that physicians screen for symptoms of cardiac dysfunction and monitor for any signs that suggest increased risk over the course of treatment. In addition, the panel indicated that careful use of stimulants is possible in the context of structural, rhythmic, or blood pressure concerns, provided that there is close medical monitoring. Finally, after much deliberation (and some highly publicized dissenting opinions), the FDA elected not to issue a black-box warning (ie, a warning in the product insert that is highlighted in a black box). Instead, a mid-level safety warning was issued, which specifically highlighted the risks of treatment in patients with structural cardiac anomalies, cardiomyopathies, heart rhythm abnormalities, coronary artery disease, or other cardiac problems. The warning also advised clinicians to conduct pretreatment screening for these conditions and, as previously advised by

the AHA, to monitor pulse and blood pressure over the course of treatment.

Commentary on the Use of Postmarketing Surveillance To Determine the Risk of Sudden Cardiac Events with Treatment

Although the reported rate of sudden cardiac death in treated individuals was lower than the rates in the general population, these figures cannot be interpreted to mean that there is not a relationship between sudden cardiac death or other untoward cardiac events and stimulant use. Complicating any interpretation are the recognition that the frequency of these events is likely greater than can be established from voluntary reporting and the fact that it is extremely difficult to obtain accurate epidemiologic data regarding the occurrence of infrequent events in a sufficiently large population to draw meaningful conclusions. It remains possible that there is a signal (ie, an association between medication treatment and the occurrence of sudden cardiovascular events). However, it is possible to detect such a signal only if it rises above the noise floor in the analysis (ie, the base rate in the population). Because this is apparently not the case, what we can say most clearly at this point is that if there is a signal, it is very small. Furthermore, the fact that approximately half the cases with catastrophic outcome were associated with some level of risk either at baseline or during treatment suggests that there is opportunity to further decrease whatever risk there might be.

Consensus Statement on the Assessment of Cardiovascular Risk

Following the publication of the *Circulation* article and the subsequent call for further review by the AAP, the AHA and AAP decided to jointly re-examine their clinical guidelines. On May 16, 2008, the 2 organizations held a joint news conference to announce their conclusions,⁹ and the consensus statement was published in the August 2008 issue of the *Journal of Developmental and Behavioral Pediatrics*.¹⁰ This statement was subsequently endorsed by the AACAP, American College of Cardiology, Children and Adults with Attention-Deficit/Hyperactivity Disorder, and National Initiative for Children's Healthcare Quality and represents the current position of the AHA, AAP, and AACAP.

The consensus recommendations are as follows:

1. It is essential to obtain a thorough history of cardiovascular function in all children with ADHD

as well as a history of cardiovascular risk in the biological family prior to the initiation of stimulant treatment. This recommendation was embedded in a larger directive by the AAP to carefully assess all children for cardiovascular risk, including those starting stimulants, with a targeted cardiac history (eg, a patient history of previously detected cardiac disease, palpitations, syncope, or seizures; a family history of sudden death in children or young adults; hypertrophic cardiomyopathy; and long QT syndrome) and a physical examination, which includes a careful cardiac examination (evidence quality: C; strength: recommendation). If the patient history, family history, or physical examination is suggestive of cardiac disease, further evaluation is advised with input from a pediatric cardiologist. This is a class I recommendation.

2. It is reasonable for a physician to consider obtaining an ECG as part of the pretreatment evaluation of children with ADHD. This decision should be made by the physician on a case-by-case basis after a review of the relevant historical information and physical examination and is not mandatory. This is a class IIa (ie, lower level) recommendation.
3. It is reasonable to use ADHD medications with caution in patients with known congenital heart disease and/or arrhythmias if these patients are stable and under the care of a pediatric cardiologist.

The AAP made the following additional comments/recommendations¹¹:

1. Given current evidence, the AAP encourages primary care and subspecialty physicians to continue currently recommended treatment for ADHD, including stimulant medications, without obtaining routine ECGs or routine subspecialty cardiology evaluations for most children before starting therapy with these medications (evidence quality: D; strength: option).
2. More research is needed on risk factors for sudden cardiac death among all children and adolescents, including those with ADHD who are treated with stimulant medications. Better methods of detecting hidden cardiac disease in children should be the focus of such research efforts.

Lessons Learned

The treatment of psychiatric disorders in children involves the efforts of many stakeholders, including caregivers, medical providers, and a variety of allied professionals. It is incumbent on the expert

oversight organizations and professional societies to thoughtfully engage in a process by which they address objective data and act prudently, rapidly, and objectively to communicate new findings and changes in practice recommendations to the general public and practitioner organizations. Coordination of data analyses and harmonization of recommendations among professional organizations are key ingredients to shaping public health policy and preserving public confidence in medical interventions.

Recent issues related to the publication and broadcast of the recommendation that pretreatment ECGs should be obtained in youth with ADHD and the subsequent reworking of this recommendation remind us of the important impact that professional organizations can have on the perception of treatment risk and safety for both professionals and the lay public. Furthermore, they underscore the need for thoughtful consultation among relevant professionals and stakeholders, remaining true to data when they are available, and being clear about the limits of data when this is the case. Despite the initial confusion generated by the public release of information regarding the AHA subcommittee review, the system proved to be self-correcting, yielding a prudent and reasonable outcome for physicians, families, and patients involved in the treatment of ADHD. Potential positive outcomes of this controversy include improvements in postmarketing surveillance, further research on risk factors for sudden cardiac death in children and adults, increased awareness on the part of clinicians treating children with ADHD, and improved methods of detecting children with hidden cardiac disease.

CLINICAL CARE IN THE CONTEXT OF GUIDELINES FOR ASSESSING CARDIAC RISK

Putting Risk into Perspective

Decades of evidence demonstrate a positive risk/benefit ratio of stimulant treatment in ADHD. Untreated ADHD continues to exact severe costs from affected individuals, including major psychosocial, occupational, and relationship difficulties, and substantial financial costs from patients and society. To highlight the detrimental effects of ADHD in adults, a survey conducted in a national sample based on more than 9000 interviews demonstrated that ADHD resulted in an average of 35 lost work days per person per year (absenteeism plus reduced

productivity on the job) among adults 18 to 44 years of age.¹² Lost work days were highest among blue-collar workers (55.8 lost days per year per person). The projected cost to the nation is 120 million lost days of work per year and \$19 billion in salary equivalents. Another study found substantial lost earning potential among individuals with ADHD.¹³ Finally, longitudinal studies of youth with ADHD have documented the increased odds of premature death due to injury from a variety of other sources, and these risks far exceed those that might possibly be associated with stimulant treatment. Thus, the risk-benefit ratio strongly supports the decision to treat.

Nonetheless, assessment and monitoring of cardiovascular risk remain a central issue in ADHD treatment. Although the occurrence of sudden catastrophic cardiac events is quite low, screening by history and physical and/or laboratory examination, when indicated, can further decrease risk and is therefore strongly recommended. Screening for cardiac risk is presumably even more important in the treatment of adults with ADHD because adults normally have higher blood pressure than children and a sizable subgroup of adults with ADHD are likely to have hypertension. Given the rapidly increasing number of adults who are currently diagnosed with ADHD and treated with stimulant and nonstimulant medications that affect catecholamine function, this issue is likely to receive further attention in the years ahead.

Talking to Patients and Families About the Cardiovascular Safety of Attention-Deficit/Hyperactivity Disorder Treatments

The FDA mandates that physicians inform patients and families about cardiovascular risks associated with medication treatments for ADHD and has developed guidelines to aid in this endeavor. It is recommended that parents tell their child's doctor if they or their child has heart defects, high blood pressure, or a family history of these problems, that doctors carefully screen for such problems before starting medication, and that the blood pressure and heart rate be checked regularly during treatment. In attempting to balance the need to advocate responsibly for treatment without unduly minimizing potential risks, it is important for doctors not only to thoroughly review information regarding possible adverse events but also to provide information regarding the types of morbidity associated with untreated ADHD and the many years of experience that the field has had in treating this condition successfully.

SUMMARY AND CONCLUSIONS

Treatments for ADHD are generally considered to be highly effective and safe, but it is important to be aware of potential problems that may accrue from treatment and to appropriately screen for risk. There is little particular concern regarding cardiovascular safety in patients who have no identifiable pretreatment or family risk factors. However, it is important to remember that there are rare patients whose risk is hidden; this fact is highlighted by occasional sudden deaths in competitive athletes who have undetected cardiac disease despite having regular medical evaluations. The risk for sudden death incurred by youth taking medication for the treatment of ADHD is probably comparable to what the risk is in youth participating in competitive athletic activities. Furthermore, although the risk is extremely low, it can be further reduced by pretreatment screening of the patient's and family's medical history, by obtaining vital signs at baseline and over the course of treatment, and by monitoring for emerging cardiac-related symptoms (eg, palpitations, syncope, and unexplained chest pain). It is hoped that future research will more precisely elucidate aspects of cardiovascular adverse effects associated with treatments for ADHD and aid in the identification of patients at highest risk. It is additionally hoped that pharmacogenomic screening will prove to be advantageous.

DISCLOSURES

Potential conflict of interest: J. Newcorn. Dr. Newcorn is a recipient of grants for research support from Eli Lilly and McNeil Pediatrics. He is a consultant and/or advisor for Eli Lilly, Novartis, McNeil, Shire, Abbott, Psychogenics, Sanofi-Aventis, Lupin, and BioBehavioral Diagnostics.

C. Donnelly. Dr Donnelly serves as a consultant to Eli Lilly and has received grant and research funds from this company. He has also served as a consultant to Johnson and Johnson and is on the speakers bureau for this company; he has received research funding from the NIMH (STAART: Citalopram for Children With Autism and Repetitive Behavior Study, #U54 MH066398-04).

REFERENCES

1. Vetter VL, Elia J, Erickson C, et al. Cardiovascular monitoring of children and adolescents with heart disease receiving medications for attention deficit/hyperactivity disorder: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young Congenital Cardiac Defects Committee and the Council on Cardiovascular Nursing. *Circulation* 2008; 117: 2407–2423.
2. Gutgesell H, Atkins D, Barst R, et al. Cardiovascular monitoring of children and adolescents receiving psychotropic drugs: a statement for healthcare professionals from the Committee on Congenital Cardiac Defects, Council on Cardiovascular Disease in the Young, American Heart Association. *Circulation* 1999; 99: 979–982.
3. Newcorn JH, Ivanov I, Sharma V, et al. Attention deficit and Disruptive Behavior Disorders. (2008) In: Tasman A, Kay J, Lieberman JA, (Eds.) *Psychiatry*, 3rd edition W.B. Saunders: Philadelphia.
4. US Food and Drug Administration. Drug Safety and Risk Management Advisory Committee report. Available at: <http://www.fda.gov/ohrms/dockets/ac/cder06.html#drugsafetyriskmgmt>. Accessed May 2008.
5. Villalba L. Safety review: sudden death with drugs used to treat ADHD. February 28, 2006. Available at: http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4210b_07_01_safetyreview.pdf. Accessed May 2008.
6. Wilens TE, Prince JB, Spencer TJ, Biederman J. Stimulants and sudden death: what is a physician to do? *Pediatrics* 2006; 118: 1215–1219.
7. Bauchner H, Stein MT. The AHA's recommendation for ECG screening in children and adolescents who take stimulant medication is premature. *Journal Watch Pediatrics and Adolescent Medicine*, April 30, 2008.
8. US Food and Drug Administration. Pediatric Advisory Committee report. Available at: <http://www.fda.gov/ohrms/dockets/ac/oc06.htm>. Accessed May 2008.
9. O'Keefe L. ECGs for all ADHD patients? AAP-AHA release joint 'clarification' on AHA recommendation. *AAP News* 2008; 29: 1.
10. American Academy of Pediatrics/American Heart Association clarification of statement on cardiovascular evaluation and monitoring of children and adolescents with heart disease receiving medications for ADHD: May 16, 2008. *J Dev Behav Pediatr* 2008; 29: 335.
11. Perrin JM, Friedman RA, Knilans TK. Cardiovascular monitoring and stimulant drugs for attention-deficit/hyperactivity disorder. *Pediatrics* 2008; 122: 451–453.
12. Kessler RC, Adler L, Ames M, et al. The prevalence and effects of adult attention deficit/hyperactivity disorder on work performance in a nationally representative sample of workers. *J Occup Environ Med* 2005; 47: 565–572.
13. Biederman J, Faraone SV, Spencer TJ, et al. Functional impairments in adults with self-reports of diagnosed ADHD: a controlled study of 1001 adults in the community. *J Clin Psychiatry* 2006; 67: 524–540.

SPECIAL FEATURE

Physical Exercise and Depression

Marije aan het Rot, PhD, Katherine A. Collins, MSW, and Heidi L. Fitterling, MPH

Mood and Anxiety Disorders Program, Department of Psychiatry, Mount Sinai School of Medicine,
New York, NY

Recently, the US Department of Health and Human Services issued its first ever physical activity guidelines, which were developed because “we clearly know enough now to recommend that all Americans . . . engage in regular physical activity to improve overall health and to reduce risk of many health problems.”¹ The current weekly recommendation (2.5 hours of moderate aerobic physical activity, 1.25 hours of vigorous aerobic physical activity, or an equivalent combination of moderate and vigorous aerobic physical activity plus additional muscle strengthening activities on 2 or more days) has increased from the 1995 recommendation issued by the American College of Sports Medicine and the Centers for Disease Control and Prevention of at least 30 minutes of moderate physical activity on most days of the week. The Mayo Clinic lists several medical benefits of regular physical activity on its Web site, including weight management, increased cardiovascular function, prevention and control of chronic diseases, and improvement in sleep.² Number 1 on its list is the positive effect of physical activity on psychological well-being.

Major depressive disorder (MDD) is characterized by periods of depressed mood and/or anhedonia (ie, loss of interest or pleasure) that last at least 2 weeks in combination with several somatic symptoms (changes in appetite, sleep, energy level, and psychomotor function) and cognitive disturbances (feelings of worthlessness or inappropriate

guilt, trouble concentrating or making decisions, and suicidal thoughts) that cause clinically significant distress or impairment in everyday functioning.³ Primary care patients often present with a lack of energy and other somatic symptoms rather than with overt psychological symptoms. Indeed, anergia is as common as depressed mood, with both symptoms affecting around 75% of depressed patients.⁴

Feeling tired and fatigued might prevent patients with MDD from engaging in physical activities, including voluntary exercise. In a 2004 study of primary care outpatients, those with severe depression had lower activity levels than those with no to moderate depression.⁵ This is an important observation, especially because exercise may reduce symptoms in patients with MDD.⁶ In this review, we summarize relevant epidemiological data on the link between depression and physical (in)activity, discuss some of the more recent exercise intervention studies in patients with MDD and in individuals reporting subclinical mood problems, and provide an overview of the biological and psychological mechanisms that may underlie the beneficial effects of exercise on mood.

DEPRESSION AND PHYSICAL ACTIVITY: EPIDEMIOLOGICAL DATA

Depression has been found to affect more than 340 million people worldwide.⁷ In the United States, it is estimated that approximately 16% of the population will meet criteria for MDD in their lifetime, with women being 1.7 times more likely than men to develop the disorder.⁸ Depression is the leading cause of disability worldwide and is predicted to be the second largest contributor to the global burden of disease by the year 2020.⁹

A negative relationship between physical activity and the symptoms and prevalence of depression has been consistently documented.^{10,11} Likewise, physical inactivity has been associated with an

Address Correspondence to:

Marije aan het Rot

Mood and Anxiety Disorders
Program, Department of
Psychiatry, Mount Sinai School
of Medicine, New York, NY
Email: marije.aanhetro@
mssm.edu

increased risk for and prevalence of depression. In 2003, Goodwin¹² published findings from the National Comorbidity Survey, a US representative sample of 8098 adults (15–54 years old). She found that regular physical activity, whether in the context of work or recreation, was associated with a significantly reduced likelihood of having current MDD. This association remained after she controlled for differences in sociodemographic characteristics, comorbid mental disorders, and self-reported physical illnesses. Similarly, in 2006, Galper and colleagues¹³ reported their findings from the Aerobics Center Longitudinal Study, an ongoing national cohort study of 5451 men and 1277 women (20–88 years old) from all 50 states, who tended to be college-educated Caucasians from middle to upper economic strata. They found an inverse relationship between physical activity and depressive symptom scores. Inactive men and women were found to be more severely depressed than their active counterparts. Furthermore, a dose-response relationship was evidenced, such that those defined as insufficiently active had significantly more depressive symptoms than those defined as sufficiently active.

In 2000, Hassmen and colleagues¹⁴ reported data on a cross-sectional study of 1856 women and 1547 men (25–64 years old) from the Finnish cardiovascular risk factor survey. Data were collected on exercise frequency and depression among other measures. They found that depression scores increased as exercise frequency decreased. Highest depression scores were found for individuals who could not exercise at all because of illness or handicap. Interestingly, they found that those individuals who reported exercising most frequently (ie, daily) had slightly higher depression scores than those who exercised 2 to 3 times or once a week. This was explained by a focus on competitiveness rather than health in this population.

The American College of Sports Medicine defines 3 types of exercise: cardiovascular training (aerobic), resistance or muscular training (anaerobic), and improvement of joint flexibility. In 2001, 6 years after the American College of Sports Medicine and the Centers for Disease Control and Prevention issued their public health recommendation on physical activity,¹⁵ the Centers for Disease Control and Prevention reported that approximately 55% of the US adult population did not meet this recommendation.¹⁶ This may account for the finding that, together with poor diet, physical inactivity is responsible for approximately 15% of all deaths in the United States, and this makes it 1 of the top 2 leading causes of mortality.¹⁷

Women as a group are not only more likely than men to have a diagnosis of MDD; they are also less likely than men to be physically active. The Centers for Disease Control and Prevention reported in 2005 that 25.9% of women participated in no physical activities at all in the month preceding the survey versus only 21.4% of men.¹⁶ Brown and colleagues¹⁸ reported the results from a prospective study of physical activity and depressive symptoms in women begun in 1996, the Australian Longitudinal Study on Women's Health. A fairly representative sample of middle-aged women (45–50 years), with overrepresentation of Australian-born, employed, and university-educated women, provided self-report data on factors affecting health and well-being at 2- to 3-year intervals. Analyses were conducted on a sample of 9207 middle-aged women who completed all 3 surveys at baseline in 1996 and at 2 follow-ups in 1998 and 2001, reporting on levels of physical activity and completing 2 measures of depressive symptoms. Results indicated that physical activity at any level is associated with decreased risk for depressive symptoms and demonstrated that increasing levels of physical activity are associated with decreasing depressive symptoms.

The association between physical activity and psychological well-being is considered to be well established. However, studies that use cross-sectional data to demonstrate a negative association between physical activity and the prevalence or severity of depression are unable to determine causality. Such an association may exist because depression has an effect on the motivation to exercise, such that people who are depressed may be less likely to engage in physical activity. Likewise, exercise may have an effect on the severity of depression, such that people who exercise are less likely to suffer from MDD. In order to establish causality, longitudinal and intervention studies are needed.

EXERCISE INTERVENTION STUDIES

In 2001, the *British Medical Journal* published a systematic review of the antidepressant effect of exercise treatment. The authors included 14 randomized controlled trials (RCTs) and reported that the effect size of exercise treatment compared to no treatment was large (Cohen's $d = 1.1$).¹⁹ In clinical practice, this can be translated into a significant exercise-induced reduction in people's scores on a depression rating scale and thus presumably into a reduction in the severity of depressive symptoms. However, most studies included in the review did

not use a dichotomous measure representing the presence or absence of an MDD diagnosis. Although symptom reduction may be considered an important goal of treatment even when symptom remission is not necessarily achieved, additional limitations of this review are that many studies were done in individuals with subclinical mood problems rather than in patients with a clear diagnosis of MDD, and few studies were conducted under adequately blinded conditions.

The idea that exercise treatment is more effective than no treatment is supported by at least 2 other systematic reviews^{20–22} as well as several narrative reviews.^{23,24} In addition, most reviews conclude that the antidepressant effect size of exercise can be comparable to that of psychotherapy and to that of pharmacotherapy.^{19,22,23,25} Moreover, the antidepressant effect of combined treatment may be larger than that of exercise alone.^{22,23,26} Given possible acute effects of exercise on mood,^{22,27–29} exercise in combination with pharmacotherapy might result in a faster onset of antidepressant action. If confirmed in a well-controlled trial, this idea could be especially important, given that antidepressant medications generally take several weeks before a clinically relevant improvement is seen. Finally, exercise interventions have also been used in combination with other lifestyle changes, such as increased bright-light exposure and vitamin supplementation.^{30–32} In general, exercise is thought to constitute a viable combination or augmentation strategy, especially because it is unlikely to interact negatively with other treatments.

Major Depressive Disorder Versus Depressed Mood

Although some researchers have argued that study findings of exercise benefit in subclinical populations with mood complaints do not necessarily generalize to patients with MDD, others believe that these studies are necessary to sufficiently prove the therapeutic benefit of exercise.⁶ Adherence to a standardized exercise regimen during an RCT may be problematic for many severely depressed patients. In addition, attempts to motivate or persuade them to participate could be considered a form of psychotherapy, thereby confounding the exercise effect. Studies in individuals with subclinical mood problems are informative, especially if one considers depressive episodes as occurring along continuums of both severity and length. The results from RCTs done in subclinical populations can certainly be extrapolated to predict the potential benefit of exercise in motivated patients with MDD. Exercise

is thought to be at least as efficacious in reducing depression symptoms in depressed patients as in healthy individuals who are simply feeling down.^{22,33} Also, exercise regimens are thought to have dropout rates that are similar to those for other treatments (eg, pharmacotherapy); in other words, motivation in patients with MDD may be a limiting factor, regardless of the intervention offered.^{21,23} Conversely, adherence to an exercise intervention may be as high as adherence to a medication intervention; this supports the use of exercise as a viable treatment strategy for MDD.

Findings of 2 Dose-Response Studies

To reduce the placebo effect caused by people's expectations of the treatment, RCTs in which both the research participants and the investigators are blinded are desirable. However, in exercise RCTs, the blinding of treatments (exercise versus no treatment, exercise versus pharmacotherapy, or exercise versus psychotherapy) has proven difficult. One way in which the expectancy effect can be reduced is direct comparison of different exercise doses in different groups of patients. Because all patients in a trial are exercising, they will all expect the benefit. For example, Dunn *et al.*³⁴ compared the weekly aerobic exercise (cardiovascular training) dose recommended by the American College of Sports Medicine with a lower dose in a group of patients with mild to moderate MDD. After 12 weeks, the public health dose was found to be effective in reducing depression symptoms at a rate comparable to other antidepressant interventions, whereas the low exercise dose was not. A similar dose-response effect was found for anaerobic exercise (resistance training) in older patients with MDD.³⁵ In this second study, high-dose resistance training not only reduced depression symptoms but also improved patients' subjective sleep quality and quality of life and increased muscle strength. Although these 2 RCTs had several limitations (eg, group differences in dropout rates and in medical care), they arguably provide the strongest evidence for the antidepressant effect of exercise at the recommended public health dose to date.

The 2 dose-response studies also show that both aerobic and anaerobic exercise can be effective in reducing symptoms of depression; this finding is consistent with earlier reports and especially relevant for clinical practice (eg, for patients who are unwilling or unable to adhere to cardiovascular training, resistance training is a viable alternative).^{22,36,37} Furthermore, the results from these 2 studies argue against the idea that social aspects of the exercise treatment (meeting with a trainer and a group setting)

can fully explain its antidepressant effects: patients in both dosing groups interacted with their trainer and/or group members at a similar frequency. This does not mean that the social interaction associated with many forms of exercise will not benefit mood. The antidepressant effect size of exercise may be larger than that of a social control intervention.³⁸ On the other hand, there is some evidence that social contact may be as efficacious in reducing core depression symptoms as trainer-assisted exercise.³⁹ However, exercise may be more likely to also reduce somatic symptoms (eg, improve sleep).³⁹

Duration of the Antidepressant Effect of Exercise

The benefit of exercise is known to be greater for longer exercise programs and/or more sessions, although even a single session may be efficacious in acutely reducing symptoms of depression.^{22,27-29} It is presently unclear how long the antidepressant effect of exercise may last beyond the end of a program. However, follow-up depression scores might remain low for at least some time.²² Moreover, individuals who start exercising during a trial may be more likely to continue exercising after the trial than nonexercisers in the trial are likely to start.⁴⁰ Sustained adherence to an exercise regimen after a trial has ended has been found to result in fewer depression symptoms when participants are followed up weeks, months, or even years later.^{40,41} Together, these data suggest that exercise treatment may have some long-lasting benefits that cannot be explained by its immediate impact on the biology and psychology underlying mood alone.

NEUROBIOLOGY OF EXERCISE

Given that the pathophysiology of MDD is both complex and still poorly understood, it is not surprising that there is no consensus on the putative biological mechanism underlying the positive effect of exercise on the regulation of mood. However, a review of the literature suggests that a number of biological systems may be involved.

Exercise, Opioids, and Cannabinoids

Perhaps the most well-known neurochemical correlate of exercise is the biological change underlying the "runner's high," which is defined as a sense of euphoria and analgesia (pain insensitivity) experienced after strenuous physical activity. The leading explanation for this phenomenon is an increase

in endogenous opioids known as endorphins.⁴² Proponents of this endorphin hypothesis cite as evidence both raised levels of endorphins measured in blood plasma and cerebrospinal fluid after running and a blunting of the runner's high after administration of naloxone, an opioid receptor antagonist.⁴³⁻⁴⁵ A recent brain imaging study in healthy long-distance runners used positron emission topography to demonstrate increased opioid activity in several frontolimbic brain regions implicated in mood regulation after 2 hours of running.⁵⁰ The magnitude of change in opioid binding correlated with the intensity of self-reported euphoria.⁵⁰

Some have questioned the validity of the endorphin hypothesis, noting, for instance, that an increase in endogenous opioid activity should produce not only euphoria and analgesia but also respiratory depression, pupillary constriction, and reduced gastrointestinal motility, effects not observed in runners.⁵¹ Dietrich and McDaniel⁵¹ offer an alternative endocannabinoid hypothesis. Endocannabinoids, like the endogenous opioids, are known to have anxiolytic and analgesic properties and to be present in elevated concentrations in the body following exercise. Hence, Dietrich and McDaniel believe an exercise-induced increase in endocannabinoid activity may also contribute to the runner's high, although they acknowledge the need for further research to determine how these systems and others interact.⁵¹

The runner's high phenomenon does not occur in every person, nor does it occur consistently or after every run. Nevertheless, the mere existence of the runner's high phenomenon evinces exercise's ability to directly improve mood. Furthermore, there is evidence from animal studies, post-mortem studies in suicide victims, and human brain imaging studies that opioid and endocannabinoid activity may be altered in patients with MDD.⁵³⁻⁶⁰ However, at this time, no coherent model of the pathophysiology of MDD incorporates dysfunction in these particular neuromodulatory systems. Thus, it is presently unclear if and how the physiological processes that precipitate the runner's high might help explain the antidepressant effects of exercise.

Exercise and the Monoamines

Other bodies of research aiming to understand the interaction between exercise and MDD are anchored by preexisting ideas about the pathophysiology of MDD. In particular, the monoamine hypothesis of depression, which originally postulated that inadequate availability of the monoamine neurotransmitters serotonin and norepinephrine

results in a depressed state, has dominated mechanistic and treatment research since the 1950s, including research investigating links between exercise and mood. Data showing that exercise increases the availability of serotonin and norepinephrine are of particular interest, given that most antidepressant medications currently approved by the US Food and Drug Administration, including the monoamine oxidase inhibitors, tricyclics, selective serotonin reuptake inhibitors, and serotonin and norepinephrine reuptake inhibitors, are known to increase serotonin and/or norepinephrine levels.

Several lines of evidence suggest that exercise stimulates brain serotonin function. In experimental animals, physical activity has been shown to increase levels of tryptophan, the amino acid precursor of serotonin, in blood plasma and in cerebrospinal fluid, to boost serotonin neurotransmission, to enhance serotonin metabolism, and to increase the production of various proteins involved in brain serotonin function.^{61–66} The idea that exercise may also stimulate serotonin metabolism in humans is largely based on a study that found increased levels of a serotonin metabolite, 5-hydroxyindoleacetic acid, following physical activity.⁶⁷ Exercise is also known to decrease plasma levels of the branched chain amino acids valine, leucine, and isoleucine.^{68–70} These amino acids compete with tryptophan for uptake into the brain. An exercise-induced increase in the ratio of tryptophan levels to levels of the branched chain amino acids may thus enhance serotonin synthesis by increasing the brain availability of its precursor.

In experimental animals, engaging in an acute exercise trial induces an immediate increase in the activity of the brain cells that produce norepinephrine.^{71–73} Chronic exercise also results in increased levels of norepinephrine and its metabolites as well as activation of tyrosine hydroxylase, an enzyme involved in the production of norepinephrine.^{62,74–78} This is consistent with norepinephrine's positive effects on mood. Although it should be noted that the norepinephrine surge seen during acute exercise tends to be followed by a depletion of norepinephrine levels, an observation consistent with the idea that exercise can be stressful and exhausting and may lead to a decreased likelihood of or increased threshold for future exercise, norepinephrine depletion is not seen with chronic exercise.

Increases in dopamine activity following exercise have also been found, though less consistently so.^{67,79} Nonetheless, a significant body of indirect evidence suggests that exercise does affect dopamine function in the brain. For example, regular exercisers

can become addicted to physical activity and may experience withdrawal symptoms when inactive.^{80,81} Also, exercise has ameliorative and even protective benefits in patients with Parkinson's disease.⁸² Dopamine is known to play a role both in the development and maintenance of addiction and in the pathophysiology of Parkinson's disease.^{83,84} Dopamine's role in MDD has largely been described with respect to the motivational problems and anhedonia that many patients describe.⁸⁵

Despite data suggesting a role for the monoamines in both MDD and the effects of exercise on mood, in recent years, scientists have increasingly recognized the inadequacy of the monoamine hypothesis of depression. Many of the currently available antidepressant medications are not at all effective in approximately one-third of individuals who suffer from MDD.⁸⁶ As a result, enthusiasm for the monoamine hypothesis has waned. At the same time, a number of supplemental theories of MDD pathophysiology have emerged. Two relevant neurochemical theories focus on a neuroendocrine feedback system consisting of the hypothalamus, pituitary, and adrenal glands [the hypothalamic-pituitary-adrenal (HPA) axis] or on the role of neurotrophic factors such as brain-derived neurotrophic factor. In addition, neuroimaging researchers in particular are concerned with the role of abnormalities in brain circuitry in MDD.

Exercise and the Hypothalamic-Pituitary-Adrenal Axis

The HPA axis, responsible for mounting the body's response to stress by coordinating the release of various stress hormones that determine the state of arousal, is known to function abnormally in the depressed state in some if not all individuals with MDD.⁸⁷ A failure of the hippocampus and hypothalamus to inhibit the release of these stress hormones is thought to result in chronic hypercortisolemia, even in the absence of external stressors. The observed effects of physical activity on the HPA axis suggest that exercise may be able to reverse or reduce a potential HPA axis abnormality. Although exercise may acutely increase levels of the stress hormones corticotropin and cortisol (which in terms of mood may mainly help induce a state of vigor), long-term exercise training appears to blunt the body's response not only to the stress of exercise but also to stress in general.^{88–90} Interestingly, the effects of acute exercise versus chronic exercise on the HPA axis parallel those on norepinephrine, and this is consistent with the interactions that are known to exist between these 2 systems.⁸⁷

Exercise and Neurotrophins

The neurotrophic hypothesis of depression has been bolstered by observations that stress impairs, whereas successful antidepressant treatment enhances, the brain's ability to generate new cells and support existing cells. On the basis of these findings, it has been argued that factors regulating the birth and health of neurons across neurotransmitter systems, rather than abnormalities in specific neurotransmitters, underlie the pathophysiology of MDD.^{91,92} As summarized in Table 1, exercise is known to increase levels of a number of neurotrophic factors in both experimental animals and humans (see van Praag⁹³ for a recent review).

Exercise and Neurocircuitry

Human neuroimaging investigations have contributed substantially to the understanding of the neurobiology of MDD. The depressed state is characterized by abnormal functional activity in various brain regions including the prefrontal cortex, amygdala, and hippocampus. In particular, a pattern of altered activity in ventral regions of the anterior cingulate prefrontal cortex has been consistently demonstrated.⁹⁴ A reversal is seen with successful antidepressant treatment.⁹⁵ Of potential direct relevance to the neurobiology underlying the antidepressant effect of exercise are findings that exercise may induce changes in anterior cingulate prefrontal cortex activity that parallel those associated with the therapeutic response to antidepressant medications.⁹⁶

that there is a positive association between physical activity and mental health. For example, the self-efficacy theory, originally proposed by Bandura in 1977,⁹⁷ postulates that confidence in one's ability to exercise is very much related to a person's actual ability to exercise. Self-efficacy has indeed been found to be strongly related to both the adoption and maintenance of an exercise program.^{98–100} Another hypothesis posits that feelings of success, control, and independence result from an increasing mastery of one's physical skills. It is suggested that exercisers gain a sense of confidence as they become more skillful and that this feeling of mastery is translated into other areas of their lives, including the management of their depressive symptoms and adherence to their exercise program.^{43,101,102} Furthermore, although social interaction alone may not fully explain the antidepressant impact of exercise (as discussed previously), social interaction may play an integral role in promoting the mood-enhancing effects of those forms of physical activity that are undertaken in a social context, such as team sports and socially supported physical activity.¹⁰³

On the other hand, in order to effectively encourage people to develop and maintain an appropriate level of physical activity, it is important to be aware of the psychological factors that may act as barriers to being physically active.¹⁰⁴ First, depression itself may pose such a barrier.¹⁰⁵ Symptoms such as loss of energy, fatigue, diminished pleasure, and feelings of worthlessness will make exercise a more difficult undertaking than it would be in the absence of such symptoms. Furthermore, psychological barriers to exercise may include fear, negative experiences with exercise in the past, and a perceived lack of knowledge about available exercise options.¹⁰⁶ There is also some evidence that depressed mood may bias people's perceptions of how strenuous exercise is.¹⁰⁷ In depressed

PSYCHOLOGY OF EXERCISE

Various psychological mechanisms have been proposed to explain, in part, the consistent finding

Table 1. Levels of Various Neurotrophins Are Increased by Physical Exercise.

Neurotrophic Factor	Function	Impact of Exercise	References
Brain-derived neurotrophic factor	Factor that promotes cell multiplication, cell survival, and synaptogenesis	<i>In humans:</i> increased levels in serum <i>In rats:</i> increased gene transcription in hippocampus (specifically the dentate gyrus)	108–111
Insulin-like growth factor 1	Hormone that activates pathways which promote cell growth and multiplication and that inhibits programmed cell death	<i>In rats:</i> increased uptake in brain cells and reduced neurogenesis when blocked	112, 113
Vascular endothelial growth factors	Platelet-derived proteins that promote angiogenesis (including new brain tissue)	<i>In humans:</i> increased levels in blood serum/blood plasma and increased gene transcription factors <i>In rats:</i> reduced neurogenesis when blocked	114–116

overweight youth, who may benefit in several ways from physical activity, high levels of parent distress and peer victimization are significant barriers to physical activity.¹⁰⁵ For women, who are particularly vulnerable to both depression and physical inactivity, the domain and social context of an exercise program have been found to be important factors influencing the effect of physical activity on mental health.¹⁰⁸ Overall, it is important for clinicians to understand that individual patients will each have their own unique set of circumstances and perceptions that may act as barriers to being physically active. Only when one knows what these are can an appropriate approach to overcoming the resistance to exercise be taken.

Finally, it is important to keep in mind that many if not all of the psychological mechanisms that underlie the effects of exercise on mood not only are interactive but also work bidirectionally. For example, while exercise may improve mood by increasing one's self-worth, mood improvement might in turn increase the likelihood that someone exercises again through its impact on self-worth. The same is likely true for the biological mechanisms that underlie the links between exercise and mood.

CONCLUSIONS AND FUTURE DIRECTIONS

Epidemiologic evidence for the positive effects of physical activity and voluntary exercise on mental health is clearly supported by data from exercise intervention studies showing that the exercise dose recommended by the Centers of Disease Control and Prevention and the American College of Sports Medicine (ie, at least 30 minutes of moderate-intensity physical activity on most days of the week) can be efficacious in reducing symptoms in patients with MDD. The somewhat higher dose recently recommended by the US Department of Health and Human Services (ie, 2.5 hours of moderate aerobic physical activity plus 2 or more days of muscle strengthening per week) will likely be more effective, even though it has not yet been officially studied. Exercise intervention studies continue to show large effect sizes, even though more recent studies have been better controlled and include larger populations. In contrast, the effect size of antidepressant pharmacotherapy in particular seems to taper off as larger, better controlled trials are being conducted.¹⁰⁹

Nevertheless, in the United States, exercise is still not a first-line or even second-line treatment option

for individuals with MDD. In contrast, in the United Kingdom, national guidelines for treating MDD explicitly list physical exercise as a particularly useful and effective intervention for patients with mild depression symptoms and recommend a structured and supervised program of 45 to 60 minutes up to 3 times a week for 10 to 12 weeks. The United Kingdom Department of Health also has an exercise referral system in place for patients consulting their doctors about depressive symptoms.¹¹⁰ The United States does not currently have national guidelines for an exercise program specifically designed for patients with MDD, even though exercise treatment has few if any adverse effects, is readily available and cost-effective, can be sustained indefinitely (unlike most other treatments for MDD), and has additional benefits for multiple aspects of physical health.¹¹¹ In particular, exercise helps improve sleep, increase muscle strength, enhance cardiovascular function, and control hypertension, hypercholesterolemia, and diabetes and can even reduce mortality.¹¹² Many of these conditions are common in MDD^{113,114}; exercise therefore has the potential to improve both mental and physical health in this patient population. Furthermore, although no RCTs of exercise as a preventative strategy have been performed to date, it has been suggested that exercise may help prevent depression in addition to treating it.¹¹⁵

It is true that not all questions surrounding the exercise-depression link have been answered at this point. Recent RCT design reports suggest that investigators are currently focusing on eliminating sources of potential bias (eg, selecting a reasonable placebo control, concealing the method of randomization, and adequate blinding of outcome assessments), increasing treatment adherence (eg, involving a social coach, providing a home-based exercise program, individualizing each patient's exercise program, and implementing a behavioral intervention to reduce barriers to exercise), and testing novel hypotheses (eg, exercise as an augmentation strategy and effects on neurobiological measures).^{116–118} Nonetheless, on the basis of the current data, exercise deserves to be considered as a viable stand-alone or adjunct treatment for depression in at least some motivated patients with MDD.

On its Web site, the Mayo Clinic also lists how to overcome common barriers to exercising.¹¹⁹ Doctors may increase their patients' motivation to voluntarily start an exercise program by encouraging them to set aside time to exercise, to make exercise a fun activity, to be creative about how to get physical activity incorporated into their routines, to set realistic goals and expectations, and to talk to their doctors about how to avoid injury. It may be especially

important to tailor exercise recommendations to individual patients.¹⁰⁴ For example, in overweight individuals, who may tend to focus on weight loss as a distant outcome, clinicians could emphasize mood improvement as an important short-term benefit of exercise. Once a patient is happier, he or she will likely be eager to pursue exercise further. This positive cycle will contribute to a healthy mind in a healthy body.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Physical Activity Guidelines Writing Group. 2008 *Physical Activity Guidelines for Americans*. Rockville, MD: US Department of Health and Human Services; 2008.
- Mayo Clinic Staff. Exercise: 7 benefits of regular physical activity. July 26, 2007. Available at: <http://www.mayoclinic.com/health/exercise/hq01676>. Accessed August 2008.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
- Tylee A, Gandhi P. The importance of somatic symptoms in depression in primary care. *Prim Care Companion J Clin Psychiatry* 2005; 7: 167–176.
- Iverson GL. Objective assessment of psychomotor retardation in primary care patients with depression. *J Behav Med* 2004; 27: 31–37.
- Salmon P. Effects of physical exercise on anxiety, depression, and sensitivity to stress: a unifying theory. *Clin Psychol Rev* 2001; 21: 33–61.
- Greden JF. The burden of recurrent depression: causes, consequences, and future prospects. *J Clin Psychiatry* 2001; 62(suppl 22): 5–9.
- Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003; 289: 3095–3105.
- Lopez AD, Murray CC. The global burden of disease, 1990–2020. *Nat Med* 1998; 4: 1241–1243.
- Ross CE, Hayes D. Exercise and psychologic well-being in the community. *Am J Epidemiol* 1988; 127: 762–771.
- Stephens T. Physical activity and mental health in the United States and Canada: evidence from four population surveys. *Prev Med* 1988; 17: 35–47.
- Goodwin RD. Association between physical activity and mental disorders among adults in the United States. *Prev Med* 2003; 36: 698–703.
- Galper DI, Trivedi MH, Barlow CE, et al. Inverse association between physical inactivity and mental health in men and women. *Med Sci Sports Exerc* 2006; 38: 173–178.
- Hassmen P, Koivula N, Uutela A. Physical exercise and psychological well-being: a population study in Finland. *Prev Med* 2000; 30: 17–25.
- Pate RR, Pratt M, Blair SN, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA* 1995; 273: 402–407.
- Trends in leisure-time physical inactivity by age, sex, and race/ethnicity—United States, 1994–2004. *MMWR Morb Mortal Wkly Rep* 2005; 54: 991–994.
- Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Correction: actual causes of death in the United States, 2000. *JAMA* 2005; 293: 293–294.
- Brown WJ, Ford JH, Burton NW, et al. Prospective study of physical activity and depressive symptoms in middle-aged women. *Am J Prev Med* 2005; 29: 265–272.
- Lawlor DA, Hopker SW. The effectiveness of exercise as an intervention in the management of depression: systematic review and meta-regression analysis of randomised controlled trials. *Br Med J* 2001; 322: 763–767.
- Craft LL, Landers DM. The effect of exercise on clinical depression and depression resulting from mental illness: a meta-analysis. *J Sport Exerc Psychol* 1998; 20: 339–357.
- Stathopoulou G, Powers MB, Berry AC, et al. Exercise interventions for mental health: a quantitative and qualitative review. *Clin Psychol Sci Pract* 2006; 13: 179–193.
- North TC, McCullagh P, Tran ZV. Effect of exercise on depression. *Exerc Sport Sci Rev* 1990; 18: 379–415.
- Martinsen EW. Physical activity and depression: clinical experience. *Acta Psychiatr Scand Suppl* 1994; 377: 23–27.
- Teychenne M, Ball K, Salmon J. Physical activity and likelihood of depression in adults: a review. *Prev Med* 2008; 46: 397–411.
- Blumenthal JA, Babyak MA, Moore KA, et al. Effects of exercise training on older patients with major depression. *Arch Intern Med* 1999; 159: 2349–2356.
- Trivedi MH, Greer TL, Grannemann BD, et al. Exercise as an augmentation strategy for treatment of major depression. *J Psychiatr Pract* 2006; 12: 205–213.
- Hoffman MD, Hoffman DR. Exercisers achieve greater acute exercise-induced mood enhancement than nonexercisers. *Arch Phys Med Rehabil* 2008; 89: 358–363.
- Lane AM, Crone-Grant D, Lane H. Mood changes following exercise. *Percept Mot Skills* 2002; 94(pt 1): 732–734.
- Bartholomew JB, Morrison D, Ciccolo JT. Effects of acute exercise on mood and well-being in patients with major depressive disorder. *Med Sci Sports Exerc* 2005; 37: 2032–2037.
- Brown MA, Goldstein-Shirley J, Robinson J, Casey S. The effects of a multi-modal intervention trial of light, exercise, and vitamins on women's mood. *Women Health* 2001; 34: 93–112.
- Leppamaki S, Partonen T, Lonnqvist J. Bright-light exposure combined with physical exercise elevates mood. *J Affect Disord* 2002; 72: 139–144.
- Leppamaki S, Haukka J, Lonnqvist J, Partonen T. Drop-out and mood improvement: a randomised controlled trial with light exposure and physical exercise. *BMC Psychiatry* 2004; 4: 22.

33. Brosse AL, Sheets ES, Lett HS, Blumenthal JA. Exercise and the treatment of clinical depression in adults: recent findings and future directions. *Sports Med* 2002; 32: 741–760.
34. Dunn AL, Trivedi MH, Kampert JB, et al. Exercise treatment for depression: efficacy and dose response. *Am J Prev Med* 2005; 28: 1–8.
35. Singh NA, Stavrinos TM, Scarbek Y, et al. A randomized controlled trial of high versus low intensity weight training versus general practitioner care for clinical depression in older adults. *J Gerontol A Biol Sci Med Sci* 2005; 60: 768–776.
36. Doyne EJ, Ossip-Klein DJ, Bowman ED, et al. Running versus weight lifting in the treatment of depression. *J Consult Clin Psychol* 1987; 55: 748–754.
37. Martinsen EW, Hoffart A, Solberg O. Comparing aerobic with nonaerobic forms of exercise in the treatment of clinical depression: a randomized trial. *Compr Psychiatry* 1989; 30: 324–331.
38. Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of progressive resistance training in depressed elders. *J Gerontol A Biol Sci Med Sci* 1997; 52: M27–M35.
39. McNeil JK, LeBlanc EM, Joyner M. The effect of exercise on depressive symptoms in the moderately depressed elderly. *Psychol Aging* 1991; 6: 487–488.
40. Singh NA, Clements KM, Singh MA. The efficacy of exercise as a long-term antidepressant in elderly subjects: a randomized, controlled trial. *J Gerontol A Biol Sci Med Sci* 2001; 56: M497–M504.
41. Craft LL, Freund KM, Culpepper L, Perna FM. Intervention study of exercise for depressive symptoms in women. *J Womens Health* 2007; 16: 1499–1509.
42. Morgan WP. Affective beneficence of vigorous physical activity. *Med Sci Sports Exerc* 1985; 17: 94–100.
43. Carr DB, Bullen BA, Skrinar GS, et al. Physical conditioning facilitates the exercise-induced secretion of beta-endorphin and beta-lipotropin in women. *N Engl J Med* 1981; 305: 560–563.
44. Gambert SR, Garthwaite TL, Pontzer CH, et al. Running elevates plasma beta-endorphin immunoreactivity and ACTH in untrained human subjects. *Proc Soc Exp Biol Med* 1981; 168: 1–4.
45. Farrell PA, Gates WK, Maksud MG, Morgan WP. Increases in plasma beta-endorphin/beta-lipotropin immunoreactivity after treadmill running in humans. *J Appl Physiol* 1982; 52: 1245–1249.
46. Janal MN, Colt EW, Clark WC, Glusman M. Pain sensitivity, mood and plasma endocrine levels in man following long-distance running: effects of naloxone. *Pain* 1984; 19: 13–25.
47. Wildmann J, Kruger A, Schmale M, et al. Increase of circulating beta-endorphin-like immunoreactivity correlates with the change in feeling of pleasantness after running. *Life Sci* 1986; 38: 997–1003.
48. Radosevich PM, Nash JA, Lacy DB, et al. Effects of low- and high-intensity exercise on plasma and cerebrospinal fluid levels of ir-beta-endorphin, ACTH, cortisol, norepinephrine and glucose in the conscious dog. *Brain Res* 1989; 498: 89–98.
49. Hoffmann P, Terenius L, Thoren P. Cerebrospinal fluid immunoreactive beta-endorphin concentration is increased by voluntary exercise in the spontaneously hypertensive rat. *Regul Pept* 1990; 28: 233–239.
50. Boecker H, Sprenger T, Spilker ME, et al. The runner's high: opioidergic mechanisms in the human brain. *Cereb Cortex* 2008; 18: 2523–2531.
51. Dietrich A, McDaniel WF. Endocannabinoids and exercise. *Br J Sports Med* 2004; 38: 536–541.
52. Sparling PB, Giuffrida A, Piomelli D, et al. Exercise activates the endocannabinoid system. *Neuroreport* 2003; 14: 2209–2211.
53. Antkiewicz-Michaluk L, Rokosz-Pelc A, Vetulani J. Effect of electroconvulsive treatment on cerebral opioid receptors in the rat: changes in delta but not mu receptors. *Naunyn Schmiedeberg Arch Pharmacol* 1984; 328: 87–89.
54. Stengaard-Pedersen K, Schou M. Opioid receptors in the brain of the rat following chronic treatment with desipramine and electroconvulsive shock. *Neuropharmacology* 1986; 25: 1365–1371.
55. Gross-Isseroff R, Dillon KA, Israeli M, Biegon A. Regionally selective increases in mu opioid receptor density in the brains of suicide victims. *Brain Res* 1990; 530: 312–316.
56. Gabilondo AM, Meana JJ, Garcia-Sevilla JA. Increased density of mu-opioid receptors in the postmortem brain of suicide victims. *Brain Res* 1995; 682: 245–250.
57. Kennedy SE, Koeppe RA, Young EA, Zubieta JK. Dysregulation of endogenous opioid emotion regulation circuitry in major depression in women. *Arch Gen Psychiatry* 2006; 63: 1199–1208.
58. Adamczyk P, Golda A, McCreary AC, et al. Activation of endocannabinoid transmission induces antidepressant-like effects in rats. *J Physiol Pharmacol* 2008; 59: 217–228.
59. Hill MN, Carrier EJ, McLaughlin RJ, et al. Regional alterations in the endocannabinoid system in an animal model of depression: effects of concurrent antidepressant treatment. *J Neurochem* 2008; 106: 2322–2336.
60. Hill MN, Miller GE, Ho WS, et al. Serum endocannabinoid content is altered in females with depressive disorders: a preliminary report. *Pharmacopsychiatry* 2008; 41: 48–53.
61. Chaouloff F, Elghozi JL, Guezennec Y, Laude D. Effects of conditioned running on plasma, liver and brain tryptophan and on brain 5-hydroxytryptamine metabolism of the rat. *Br J Pharmacol* 1985; 86: 33–41.
62. Chaouloff F. Physical exercise and brain monoamines: a review. *Acta Physiol Scand* 1989; 137: 1–13.
63. Meeusen R, De Meirleir K. Exercise and brain neurotransmission. *Sports Med* 1995; 20: 160–188.
64. Wilson WM, Marsden CA. In vivo measurement of extracellular serotonin in the ventral hippocampus during treadmill running. *Behav Pharmacol* 1996; 7: 101–104.
65. Gomez-Merino D, Bequet F, Berthelot M, et al. Site-dependent effects of an acute intensive exercise on extracellular 5-HT and 5-HIAA levels in rat brain. *Neurosci Lett* 2001; 301: 143–146.
66. Greenwood BN, Foley TE, Day HE, et al. Wheel running alters serotonin (5-HT) transporter, 5-HT1A, 5-HT1B, and alpha 1b-adrenergic receptor mRNA in the rat raphe nuclei. *Biol Psychiatry* 2005; 57: 559–568.
67. Post RM, Kotin J, Goodwin FK, Gordon EK. Psychomotor activity and cerebrospinal fluid amine

- metabolites in affective illness. *Am J Psychiatry* 1973; 130: 67–72.
68. Pardridge WM. Blood-brain barrier transport of nutrients. *Nutr Rev* 1986; 44(suppl): 15–25.
 69. Davis JM, Alderson NL, Welsh RS. Serotonin and central nervous system fatigue: nutritional considerations. *Am J Clin Nutr* 2000; 72(suppl): 573S–578S.
 70. Blomstrand E. Amino acids and central fatigue. *Amino Acids* 2001; 20: 25–34.
 71. Barchas JD, Freedman DX. Brain amines: response to physiological stress. *Biochem Pharmacol* 1963; 12: 1232–1235.
 72. Stone EA. Accumulation and metabolism of norepinephrine in rat hypothalamus after exhaustive stress. *J Neurochem* 1973; 21: 589–601.
 73. Glavin GB. Stress and brain noradrenaline: a review. *Neurosci Biobehav Rev* 1985; 9: 233–243.
 74. Dishman RK. Brain monoamines, exercise, and behavioral stress: animal models. *Med Sci Sports Exerc* 1997; 29: 63–74.
 75. Soares J, Holmes PV, Renner KJ, et al. Brain noradrenergic responses to footshock after chronic activity-wheel running. *Behav Neurosci* 1999; 113: 558–566.
 76. Dishman RK, Renner KJ, White-Welkley JE, et al. Treadmill exercise training augments brain norepinephrine response to familiar and novel stress. *Brain Res Bull* 2000; 52: 337–342.
 77. O'Neal HA, Van Hoomissen JD, Holmes PV, Dishman RK. Prepro-galanin messenger RNA levels are increased in rat locus coeruleus after treadmill exercise training. *Neurosci Lett* 2001; 299: 69–72.
 78. Van Hoomissen JD, Holmes PV, Zellner AS, et al. Effects of beta-adrenoreceptor blockade during chronic exercise on contextual fear conditioning and mRNA for galanin and brain-derived neurotrophic factor. *Behav Neurosci* 2004; 118: 1378–1390.
 79. Bliss EL, Ailion J. Relationship of stress and activity to brain dopamine and homovanillic acid. *Life Sci* 1971; 10: 1161–1169.
 80. Glass JM, Lyden AK, Petzke F, et al. The effect of brief exercise cessation on pain, fatigue, and mood symptom development in healthy, fit individuals. *J Psychosom Res* 2004; 57: 391–398.
 81. Berlin AA, Kop WJ, Deuster PA. Depressive mood symptoms and fatigue after exercise withdrawal: the potential role of decreased fitness. *Psychosom Med* 2006; 68: 224–230.
 82. Herman T, Giladi N, Gruendlinger L, Hausdorff JM. Six weeks of intensive treadmill training improves gait and quality of life in patients with Parkinson's disease: a pilot study. *Arch Phys Med Rehabil* 2007; 88: 1154–1158.
 83. Goodman A. Neurobiology of addiction. An integrative review. *Biochem Pharmacol* 2008; 75: 266–322.
 84. Galvan A, Wichmann T. Pathophysiology of Parkinsonism. *Clin Neurophysiol* 2008; 119: 1459–1474.
 85. Dunlop BW, Nemeroff CB. The role of dopamine in the pathophysiology of depression. *Arch Gen Psychiatry* 2007; 64: 327–337.
 86. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006; 163: 1905–1917.
 87. Dunn AL, Dishman RK. Exercise and the neurobiology of depression. *Exerc Sport Sci Rev* 1991; 19: 41–48.
 88. Sutton JR, Young JD, Lazarus L, et al. The hormonal response to physical exercise. *Australas Ann Med* 1969; 18: 84–90.
 89. Luger A, Deuster PA, Kyle SB, et al. Acute hypothalamic-pituitary-adrenal responses to the stress of treadmill exercise. Physiologic adaptations to physical training. *N Engl J Med* 1987; 316: 1309–1315.
 90. Stranahan AM, Lee K, Mattson MP. Central mechanisms of HPA axis regulation by voluntary exercise. *Neuromolecular Med* 2008; 10: 118–127.
 91. Duman RS, Heninger GR, Nestler EJ. A molecular and cellular theory of depression. *Arch Gen Psychiatry* 1997; 54: 597–606.
 92. Duman RS, Monteggia LM. A neurotrophic model for stress-related mood disorders. *Biol Psychiatry* 2006; 59: 1116–1127.
 93. van Praag H. Neurogenesis and exercise: past and future directions. *Neuromolecular Med* 2008; 10: 128–140.
 94. Stone EA, Lin Y, Quartermain D. A final common pathway for depression? Progress toward a general conceptual framework. *Neurosci Biobehav Rev* 2008; 32: 508–524.
 95. Dougherty DD, Rauch SL. Brain correlates of antidepressant treatment outcome from neuroimaging studies in depression. *Psychiatr Clin N Am* 2007; 30: 91–103.
 96. Prakash RS, Snook EM, Erickson KI, et al. Cardiorespiratory fitness: a predictor of cortical plasticity in multiple sclerosis. *Neuroimage* 2007; 34: 1238–1244.
 97. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev* 1977; 84: 191–215.
 98. Sallis JF, Haskell WL, Fortmann SP, et al. Predictors of adoption and maintenance of physical activity in a community sample. *Prev Med* 1986; 15: 331–341.
 99. Marcus BH. Exercise behavior and strategies for intervention. *Res Q Exerc Sport* 1995; 66: 319–323.
 100. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. *Res Q Exerc Sport* 1992; 63: 60–66.
 101. Greist JH, Klein MH, Eischens RR, et al. Running as treatment for depression. *Compr Psychiatry* 1979; 20: 41–54.
 102. Paluska SA, D'Amico FJ. The comfort of family practice residents with health care of patients of the opposite gender. *Fam Med* 2000; 32: 612–617.
 103. Street G, James R, Cutt H. The relationship between organised physical recreation and mental health. *Health Promot J Austr* 2007; 18: 236–239.
 104. Seime RJ, Vickers KS. The challenges of treating depression with exercise: from evidence to practice. *Clin Psychol Sci Pract* 2006; 13: 194–197.
 105. Gray WN, Janicke DM, Ingerski LM, Silverstein JH. The impact of peer victimization, parent distress and child depression on barrier formation and physical activity in overweight youth. *J Dev Behav Pediatr* 2008; 29: 26–33.
 106. Rosqvist E, Heikkinen E, Lyyra TM, et al. Factors affecting the increased risk of physical inactivity among older people with depressive symptoms. *Scand J Med Sci Sports* 2008;

- doi: 10.1111/j.1600-0838.2008.00798.x. Available at: <http://www3.interscience.wiley.com>
107. Morgan WP. Psychological components of effort sense. *Med Sci Sports Exerc* 1994; 26: 1071–1077.
 108. Teychenne M, Ball K, Salmon J. Associations between physical activity and depressive symptoms in women. *Int J Behav Nutr Phys Act* 2008; 5: 27.
 109. Otto MW, Church TS, Craft LL, et al. Exercise for mood and anxiety disorders. *J Clin Psychiatry* 2007; 68: 669–685.
 110. *Exercise Referral Systems: A National Quality Assurance Framework*. London, England: Department of Health; 2001.
 111. Fox KR, Boutcher SH, Faulkner G, Biddle SJH. The case for exercise in the promotion of mental health and psychological well-being. In: Biddle SJH, Fox KR, Boutcher SH, eds. *Physical Activity and Psychological Well-Being*. London, England: Routledge; 2000; 1–9.
 112. Mora S, Redberg RF, Cui Y, et al. Ability of exercise testing to predict cardiovascular and all-cause death in asymptomatic women: a 20-year follow-up of the lipid research clinics prevalence study. *JAMA* 2003; 290: 1600–1607.
 113. van der Kooy K, van Hout H, Marwijk H, et al. Depression and the risk for cardiovascular diseases: systematic review and meta analysis. *Int J Geriatr Psychiatry* 2007; 22: 613–626.
 114. Carnethon MR, Biggs ML, Barzilay JI, et al. Longitudinal association between depressive symptoms and incident type 2 diabetes mellitus in older adults: the Cardiovascular Health Study. *Arch Intern Med* 2007; 167: 802–807.
 115. Southwick SM, Vythilingam M, Charney DS. The psychobiology of depression and resilience to stress: implications for prevention and treatment. *Annu Rev Clin Psychol* 2005; 1: 255–291.
 116. Trivedi MH, Greer TL, Grannemann BD, et al. TREAD: Treatment with Exercise Augmentation for Depression: study rationale and design. *Clin Trials* 2006; 3: 291–305.
 117. Krogh J, Petersen L, Timmermann M, et al. Design paper: the DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of aerobic versus non-aerobic versus relaxation training for patients with light to moderate depression. *Contemp Clin Trials* 2007; 28: 79–89.
 118. Kerse N, Falloon K, Moyes S, et al. DeLLITE depression in late life: an intervention trial of exercise. Design and recruitment of a randomised controlled trial. *BMC Geriatr* 2008; 8: 12.
 119. Mayo Clinic Staff. Barriers to fitness: overcoming common problems. February 22, 2007. Available at: http://www.mayoclinic.com/health/fitness/sm00085_d. Accessed August 2008.

**HINARI PHASE I COUNTRIES**

Free access to more than 100 biomedical journals published by Wiley

**HINARI PHASE II COUNTRIES**

Nominal cost access to more than 100 biomedical journals published by Wiley

Developing World Access to Research

“perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries.”

Former WHO Director-General Dr Gro Harlem Brundtland (July, 2001)

Wiley, together with the World Health Organisation and fellow Scientific, Technical and Medical publishers, is proud to play a role in furthering global dissemination of biomedical information.

Through the Health InterNetwork Access to Research Initiative (HINARI) www.healthinternetwork.org, free or nominal cost online access to more than 2,200 biomedical journals from 42 publishers is provided to more than 1,000 institutions in approximately 100 of the world's lowest income countries.

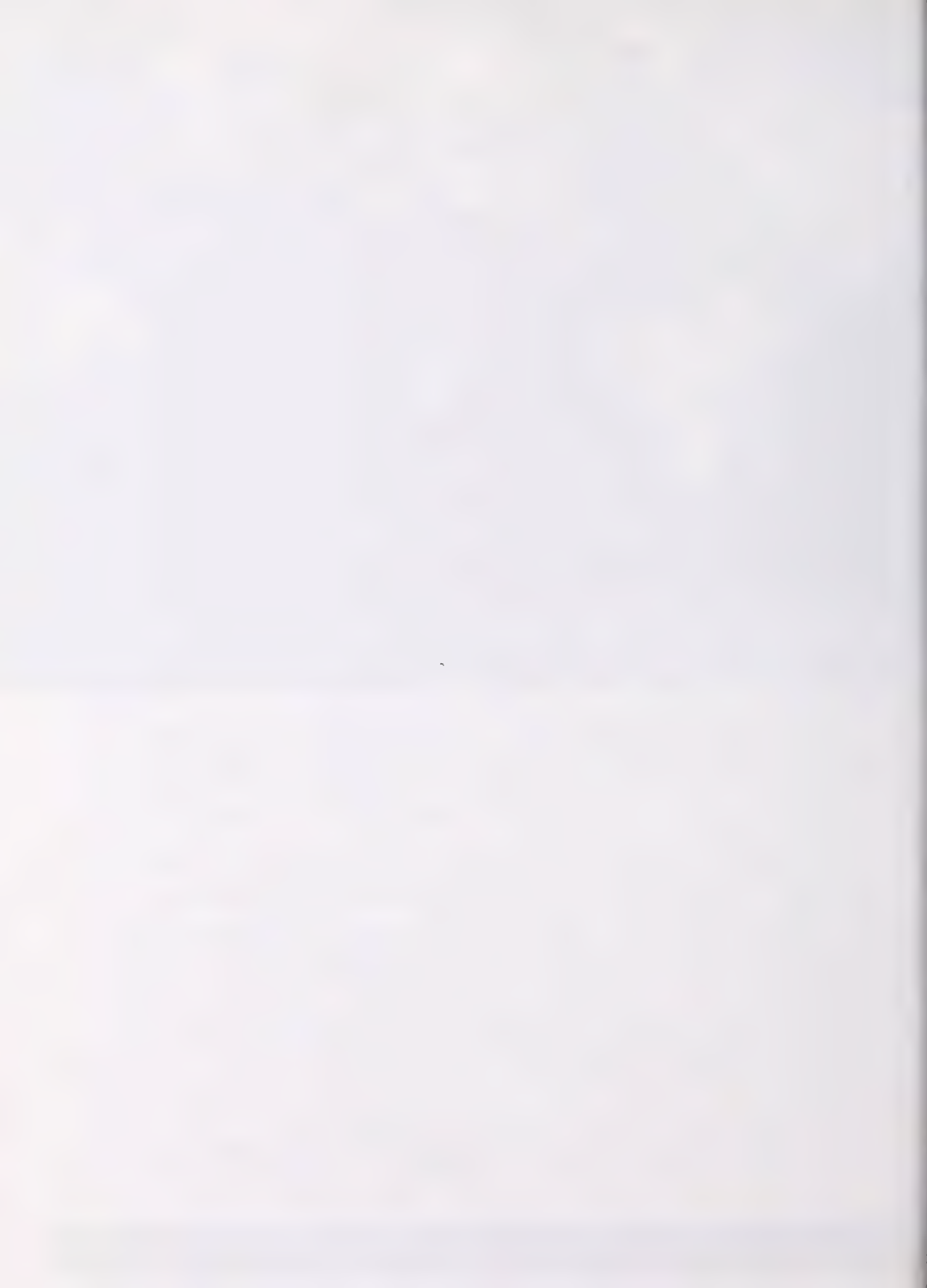
Wiley is one of the six founding publishers who led this initiative with the WHO. We offer more than 100 of our leading biomedical journals through the HINARI program; including many leading journals published with our society partners.

HINARI
Access to Research



WILEY
InterScience®
DISCOVER SOMETHING GREAT
www.interscience.wiley.com

Wiley – helping to make a world of difference.





Giant
Resources...
Purebred
Content

Wiley InterScience® Reference Works

Sample Title List

Encyclopedia of Life Sciences
Encyclopedia of Statistical Sciences
Kirk-Othmer Encyclopedia of Chemical Technology
Sax's Dangerous Properties of Industrial Materials
Ullmann's Encyclopedia of Industrial Chemistry
Van Nostrand's Encyclopedia of Chemistry

Capabilities

Full text HTML and PDF (for some)
Reference linking
Citation download

Access

Single article or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Today's sophisticated researcher demands fast and convenient online access to the giants of scientific reference.

From Kirk-Othmer to the *Encyclopedia of Life Sciences*,

Wiley InterScience® Reference Works deliver the biggest names in scientific research literature to your desktop, right when you need them. Encompassing subjects ranging from chemistry, life sciences, and medicine to psychology, engineering, and physics, these multi-volume encyclopedias, handbooks, and dictionaries are enhanced by powerful functionality that accelerates your path to discovery. With **Wiley InterScience**, you can be sure to **Discover Something Great.**



WILEY
InterScience®
DISCOVER SOMETHING GREAT



1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

Statement of Ownership, Management, and Circulation
(All Periodicals Publications Except Requester Publications)

1 Publication Title	2 Publication Number	3 Filing Date
Mount Sinai Journal of Medicine	2 8 4 8 6 0	10/1/08
4 Issue Frequency	5 Number of Issues Published Annually	6 Annual Subscription Price
6x/s/yr in April, June, Aug, Sept, Nov, Dec	6	\$95 00
7 Complete Mailing Address of Known Office of Publication (Not printer) Street, City, County, State and ZIP+4 [®]		8 Contact Person
Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030		E. Sandichen
9 Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)		Telephone (include area code)
		(201) 748-6346

Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank).
 Publisher (Name and complete mailing address)

Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

Editor (Name and complete mailing address)

Penny A. Asbell, Dept. of Ophthalmology, One Gustave L. Levy Place, Box 1183, New York, NY 10029

Managing Editor (Name and complete mailing address)

10. Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)

Full Name	Complete Mailing Address
Mount Sinai School of Medicine	1 Gustave Levy Place
	New York, NY 10029

11 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box ☒ None

[illegible]

12 Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes
☐ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

PS Form 3526, September 2006 (Page 1 of 3 (Instructions Page 3)) PSN 7530-01-000-9931 **PRIVACY NOTICE** See our privacy policy on www.usps.com

13 Publication Title	14 Issue Date for Circulation Data
Mount Sinai Journal of Medicine	August 2008

15 Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
-------------------------------------	--	---

a. Total Number of Copies (Net press run)	7228	6217
---	------	------

(1)	Mailed Outside-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	4321	2760
-----	--	------	------

b Paid Circulation 6. Mail	12. Mailed In-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	0	0
----------------------------	---	---	---

Outside the Mail)	(3) Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®	0	0
----------------------	--	---	---

(4) Paid Distribution by Other Classes of Mail Through the USPS (e.g. First-Class Mail®)	0	0
--	---	---

c. Total Paid Distribution (Sum of 15b (1), (2), (3), and (4))	4321	2760
--	------	------

(1) Free or Nominal Rate Outside-County Copies Included on PS Form 3541	353	353
--	-----	-----

d. Free or Nominal Rate Distribution	(2) Free or Nominal Rate In-County Copies Included on PS Form 3541	0	0
--------------------------------------	--	---	---

3.	Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g. First-Class Mail)	0	0
----	--	---	---

14.	Free or Nominal Rate Distribution Outside the Mail (Carriers or other means)	0	0
-----	---	---	---

e. Total Free or Nominal Rate Distribution (Sum of 15d (1), (2), (3) and (4))	353	353
---	-----	-----

f. Total Distribution (Sum of 15c and 15e)	4674	3113
--	------	------

g. Copies not Distributed (See Instructions to Publishers #4 (page #3))	2554	3104
h. Total Sum of 15 Lines a-g	7028	6217

Percent Paid	92.45	88.66
--------------	-------	-------

15 Publication of Statement of Ownership

☒ If the publication is a general publication, publication of this statement is required. Will be printed in the December 2008 issue of this publication

☐ Publication not required

7. Signature and Title of Editor, Publisher, Business Manager, or Owner	Date
	10/1/09

Elizabeth Konkle, Associate Financial Manager *elkonkle* 10/1/08

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 76, NO. 3
MAY/JUNE 2009

Issue Theme
**Type 2 Diabetes Mellitus:
New Evidence,
More Controversy**

Theme Editor
Dina Green

A Journal of Translational and Personalized Medicine



Glycemic Targets for Patients with Type 2 Diabetes Mellitus
Balancing Risk and Benefit with Oral Hypoglycemic Drugs
Targeting the Incretin System in Type 2 Diabetes Mellitus
Emerging Concepts in the Pathophysiology of Type 2 Diabetes Mellitus
Nutritional Strategies in Type 2 Diabetes Mellitus
Management of Type 2 Diabetes Mellitus and Gestational Diabetes in Pregnancy
Role of Surgery in Management of Type 2 Diabetes Mellitus

Special Features

Testing for Thrombophilia in Patients with Venous Thromboembolism: Which Tests for Whom and When
EBM on Call: Does Presenting Hyperglycemia Impact Prognosis in Non-diabetics with Acute MI?
Dilemmas in Surgery: Medical Ethics Education in the Surgery Rotation

 **WILEY-BLACKWELL**

**Newly
Relaunched!**



**Keep Abreast of the Latest Trends in
Translational and Personalized Medicine...
Enter Your Subscription Today!**

MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Editor-in-Chief:

Penny A. Asbell, MD, FACS, MBA

Mount Sinai School of Medicine, New York, NY

2008 Themes and Guest Editors

Quality of Healthcare and Racial
Health Disparities (Mark Chassin)

9/11 World Trade Center Disaster:
Past and Future (Robin Herbert,
Jacqueline Moline)

Anxiety and Depression:
Leading Edge of Therapy
(Dennis Charney, Mathew Sanjay)

Diabetes Type 1:
New Light on Etiology,
New Approaches to Management
(Derek Leroith)

Hospital Medicine:
An Ascending New Field
(Andrew Dunn)

Advanced Lung Disease:
Meeting the Challenges
(Maria Padilla)

Continuously published since 1934 and included in *Index Medicus* from the very first volume, the **Mount Sinai Journal of Medicine (MSJM)** has a distinguished career as a general medical journal for a worldwide audience.

Building upon this tradition of education across medical specialties, **MSJM** was re-launched in 2007 as a journal of translational and personalized medicine, two disciplines in the forefront of modern healthcare. **MSJM** features articles on the evolving nature of clinical care, its close linkage to bio-medical research and the resulting opportunities for enhanced patient care.

Dedicated to improving communication among all participants in medicine today, **MSJM** is an excellent source of up-to-date medical information for healthcare clinicians, educators, researchers and policy makers worldwide. **MSJM** will inspire the creative thinking needed to create new ways to understand, prevent, and treat human disease.

Each issue includes several quality, in-depth review articles related to a single theme and several others on subjects of diverse clinical interest. All articles are written by renowned authorities in their fields, from both Mount Sinai and elsewhere, under the supervision of Guest Editors.



WILEY-BLACKWELL

Enter Your 2009 Subscription Today!



Your 2009 subscription will include the six 2009 print issues of the **Mount Sinai Journal of Medicine** (Volume 76), plus concurrent full-text online access to these six issues as they are posted on Wiley's Internet publishing platform.

Two Ways for Individuals to Subscribe:

1. Mount Sinai Alumna/Alumnus:
\$40.00 (Worldwide)
2. Personal Subscription:
\$95.00 (U.S./Canada/Mexico)
\$120.00 (Outside North America)

Complimentary Online Access for Institutions!

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

PERSONAL SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 personal subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$95.00 (U.S./Canada/Mexico) ☐ \$120.00 (Outside North America)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

Method of Payment (Canada residents please add GST. Personal rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **John Wiley & Sons, Inc.** Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA

Card # _____ Exp. Date _____

Signature _____

(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM0909

MSJM MEMBER SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 Alumna/Alumnus Subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$40.00 (Worldwide)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

METHOD OF PAYMENT (Alumna/Alumnus rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **The Mount Sinai School of Medicine**. Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA ☐ Other

Card # _____ Exp. Date _____

Signature _____

(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM09

COMPLIMENTARY ONLINE ACCESS FOR INSTITUTIONS

Mount Sinai Journal of Medicine

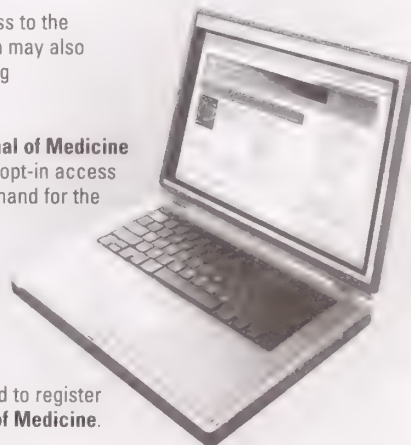
A Journal of Translational and Personalized Medicine

2009. Volume 76. 6 issues per year. Print ISSN: 0027-2507. Online ISSN: 1931-7581.

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

Institutional online and print pricing for the **Mount Sinai Journal of Medicine** will be set in 2009, subsequent to introductory complimentary opt-in access during 2008 and 2009. Pricing will be based on the overall demand for the Journal and on the size and type of the subscribing institution (e.g. academic, government, hospital, or corporate). Institutions that receive complimentary online access during 2008 and 2009 will have archival access in subsequent years according to Wiley's licensing terms and conditions.

Ask your librarian to e-mail optinaccess@wiley.co.uk or visit www.interscience.wiley.com/msjregister for more details and to register for complimentary online access to the **Mount Sinai Journal of Medicine**.





NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS MAIL

PERMIT NO. 2277

HOBOKEN, NJ

POSTAGE WILL BE PAID BY ADDRESSEE

JOHN WILEY & SONS INC
ATTN: PETER JANSIC, MAIL STOP 8-02
111 RIVER STREET
HOBOKEN, NJ 07030-9441



PLACE
POSTAGE
STAMP
HERE

INES SANTOS
MOUNT SINAI JOURNAL OF MEDICINE
ONE GUSTAV L. LEVY PLACE BOX 1094
NEW YORK, NY 10029-6574

**Library
Routing
Card**

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis S. Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Theme Editor

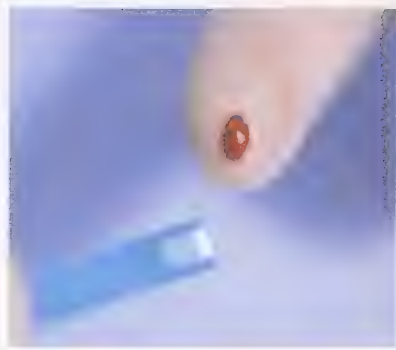
Dina E. Green MD
Assistant Professor of Medicine
Division of Endocrinology
Mount Sinai School of Medicine

Managing Editor

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com



Testing for Blood Glucose: New Evidence, More Controversy

Copyright © Photo Researchers, Inc.

Themes for future issues of *MSJM*, each shown with Guest Editor(s)

- Zoonotic infections (Stephanie Factor)
- Medical education (Lisa Coplit)
- Alzheimer's disease (Nikos Robakis)
- Endovascular surgery (Michael Marin)
- Obstetrics and gynecology (Rhoda Sperling)
- Drug discovery in academic setting (Ravi Iyengar)
- Parkinson's disease (Michele Tagliatti)

©2009 Mount Sinai School of Medicine. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: \$95.00 in North America, \$120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS

VOL 76, NO 3

Theme Introduction

- 215** Type 2 Diabetes Mellitus: New Evidence, More Controversy
Dina Green

Theme Articles

- 216** Emerging Concepts in the Pathophysiology of Type 2 Diabetes Mellitus
Prasanth N. Surampudi, Jennifer John-Kalarickal, Vivian A. Fonseca
- 227** Glycemic Targets for Patients with Type 2 Diabetes Mellitus
Ole-Petter R. Hamnvik, Graham T. McMahon
- 234** Balancing Risk and Benefit with Oral Hypoglycemic Drugs
Ole-Petter R. Hamnvik, Graham T. McMahon
- 244** Targeting the Incretin System in Type 2 Diabetes Mellitus
Matthew Potenza, Elliot J. Rayfield
- 257** Nutritional Strategies in Type 2 Diabetes Mellitus
Nichola Davis, Bernice Forbes, Judith Wylie-Rosett
- 269** Management of Type 2 Diabetes and Gestational Diabetes in Pregnancy
Tara Boinpally, Lois Jovanovič
- 281** Role of Surgery in Management of Type 2 Diabetes Mellitus
Daniel M. Herron, Winnie Tong

Feature Departments

- 294** Evidence-based Medicine [Editors: Thomas McGinn, Deborah Korenstein]
Just a spoonful of sugar: Does Presenting Hyperglycemia Impact Prognosis in Non-diabetics with Acute Myocardial Infarction?
Laurie Edelman, Thomas McGinn, Deborah Korenstein
- 297** Ethics [Editor: Rosamond Rhodes]
Dilemmas in Surgery: Medical Ethics Education in the Surgery Rotation
Nada Gligorov, Pippa Newell, Jason Altilio, M. Phil., Mike Collins, M. Phil., Amanda Favia, Leah Rosenberg, Rosamond Rhodes

Special Feature

- 303** Testing for Thrombophilia: Pitfalls, Limitations, and Marginal Impact on Treatment Duration Recommendations
Richard H. White and Robert C. Gosselin



HINARI PHASE I COUNTRIES

Free access to more than 100 biomedical journals published by Wiley

HINARI PHASE II COUNTRIES

Nominal cost access to more than 100 biomedical journals published by Wiley

Developing World Access to Research

“perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries.”

Former WHO Director-General Dr Gro Harlem Brundtland (July, 2001)

Wiley, together with the World Health Organisation and fellow Scientific, Technical and Medical publishers, is proud to play a role in furthering global dissemination of biomedical information.

Through the Health InterNetwork Access to Research Initiative (HINARI) www.healthinternetwork.org, free or nominal cost online access to more than 2,200 biomedical journals from 42 publishers is provided to more than 1,000 institutions in approximately 100 of the world's lowest income countries.

Wiley is one of the six founding publishers who led this initiative with the WHO. We offer more than 100 of our leading biomedical journals through the HINARI program; including many leading journals published with our society partners.

HINARI
Access to Research



WILEY
InterScience®
DISCOVER SOMETHING GREAT
www.interscience.wiley.com

Wiley – helping to make a world of difference.

THEME INTRODUCTION

Type 2 Diabetes Mellitus: New Evidence, More Controversy

Dina Green, MD

Division of Endocrinology, Mount Sinai School of Medicine
New York, NY

The theme of this issue of the *Mount Sinai Journal of Medicine* is "New Evidence, More Controversy." Over the past 20 years, there has been an accumulating body of evidence that treatment of type 2 diabetes with meticulous blood glucose control averts many of the dreaded microvascular and macrovascular complications of the disease. As both the prevalence of type 2 diabetes and the demand for treatment have increased, many questions have arisen. These questions largely revolve around treatment strategies for diabetes and the precise goals of treatment.

This issue opens with a description of the pathophysiology of type 2 diabetes, with a particular focus on the incretin hormones and their role in potentiating insulin action. The incretin system has recently been harnessed to produce an entirely novel class of antidiabetic agents: the incretin mimetics and dipeptidyl peptidase IV inhibitors. The mechanism of action and clinical utility of these drugs are discussed in detail in another review. Also of interest is a discussion of other oral antidiabetic agents, including the thiazolidinediones, whose side effect profile has recently come under fire. The highlight of the issue is an elegant discussion of the recent literature that asks

what constitutes an optimal level of glycemic control and what degree of glycemic control is needed to avert long-term complications. Also of interest are reviews of the role of bariatric surgery, particularly as a long-term solution to diabetes, the role of nutritional management in diabetes, and the rising prevalence of diabetes in pregnancy.

It is humbling to remember that only 90 years ago Frederick Banting and Charles Best performed early translational research when they successfully treated type 1 diabetes with insulin and transformed the natural history of the disease. In this issue, we also highlight the importance in type 2 diabetes of translational medicine, that is, the concept of applying principles of basic science to further our understanding of disease processes and treatment strategies. We can only hope that the work discussed here and other discoveries emerging from the laboratory will similarly transform our understanding of and treatment options for type 2 diabetes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

Address Correspondence to:

Dina Green

Division of Endocrinology
Mount Sinai School of Medicine
New York, NY

Email: dina.green@mountsinai.org

Emerging Concepts in The Pathophysiology of Type 2 Diabetes Mellitus

Prasanth N. Surampudi, MD, Jennifer John-Kalarickal, MD, and Vivian A. Fonseca, MD MRCP

Tulane University Medical Center, New Orleans, LA

ABSTRACT

Type 2 diabetes mellitus is a multifactorial metabolic disorder. It is characterized by chronic hyperglycemia, insulin resistance, and a relative insulin secretion defect. The prevalence of type 2 diabetes mellitus has risen worldwide in large part because of an increase in obesity and sedentary lifestyles. The underlying pathophysiology and complications of type 2 diabetes mellitus are still being elucidated. Recent advances in diabetes research have helped us to gain a better understanding about insulin resistance and insulin secretion defects. The evolving understanding about the influence of the incretin effect, insulin signal transduction, adipose tissue, intra-islet cell communication, and inflammation is changing the way in which we view type 2 diabetes mellitus. This new understanding will eventually provide us with new treatment approaches to help patients who have type 2 diabetes mellitus. This article gives a review of the current and emerging concepts of the pathophysiology of type 2 diabetes mellitus. *Mt Sinai J Med* 76:216–226, 2009. © 2009 Mount Sinai School of Medicine

Key Words: β -cell dysfunction, hepatic glucose output, incretin effect, inflammation, insulin resistance, type 2 diabetes mellitus.

Type 2 diabetes mellitus (T2DM) affects over 100 million people worldwide. Impaired glucose

tolerance (IGT) and subsequently T2DM are characterized by a relative deficiency in insulin due to the interplay between peripheral insulin resistance (IR) and insulin-secretory defects. Multiple pathophysiological impairments frequently coexist in the same patient, and it is often unclear which abnormality is the primary cause of the hyperglycemia. An improved understanding of the pathophysiology has helped to shed light on new treatment targets for T2DM. This article reviews emerging concepts in the development of T2DM.

ETIOLOGY

T2DM is a multifactorial metabolic disorder with an initial clinical manifestation of elevated blood sugars. Subjects with T2DM have elevated fasting and postprandial plasma glucose levels. The progression from normal glucose tolerance (NGT) to IGT and finally to T2DM occurs in stages. The progression is due to the interplay between IR and defects in insulin secretion (Figure 1).¹

Subjects release insulin from pancreatic β -cells to maintain normoglycemia. Individuals have various levels of biological response to insulin. In most populations, those who develop IGT and subsequently T2DM have a degree of IR even when fasting plasma glucose and glucose tolerance² are normal.^{2,3} At an early stage in the disease process, fasting insulin levels and glucose-stimulated increases in insulin responses by β -cells remain sufficient to maintain NGT. Subjects with IGT typically have increased fasting and postprandial insulin levels that do not fully compensate for IR.³ Eventually, compensation fails in some subjects because islet β -cell function declines, leading to a relative deficiency of insulin secretion (Figure 1).

The progressive loss of β -cell function and increased IR may be due to a number of factors

Address Correspondence to:

Prasanth N. Surampudi
Tulane University Medical
Center
New Orleans, LA
Email: psurampud@tulane.edu

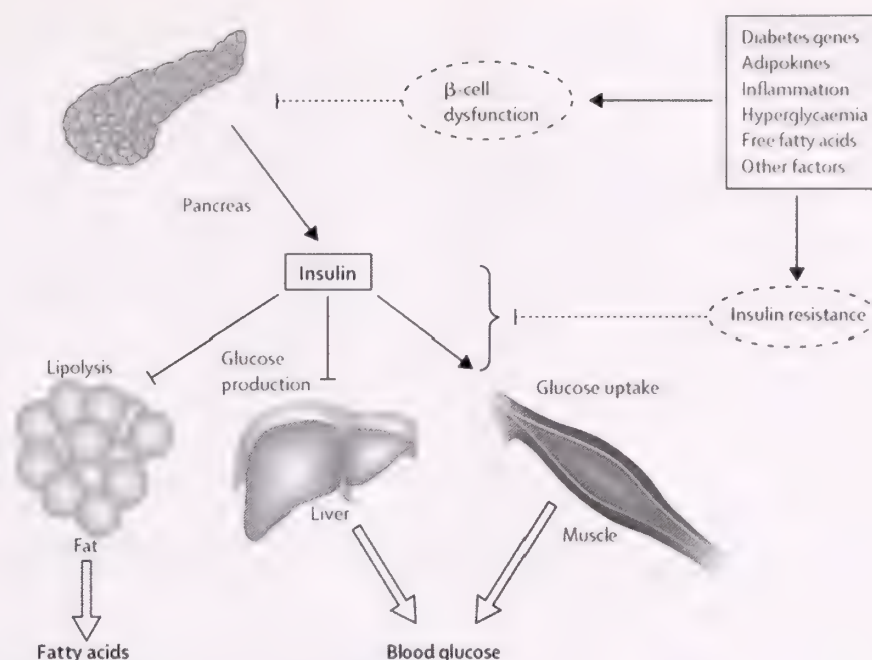


Fig 1. Abnormalities in type 2 diabetes that contribute to hyperglycemia. Insulin secretion from the pancreas normally reduces glucose output by the liver, enhances glucose uptake by skeletal muscle, and suppresses fatty acid release from fat tissue. Decreased insulin secretion will reduce insulin signaling in its target tissues. Insulin resistance affects the action of insulin in each of the major target tissues, leading to increased circulating fatty acids and the hyperglycemia of diabetes. In turn, the raised concentrations of glucose and fatty acids in the bloodstream will feed back to worsen both insulin secretion and insulin resistance (From Stumvoll, M; Goldstein, B; van Haeften, T. *Lancet*, 365:1333–1346, 2005.). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

such as genetic abnormalities and acquired defects. Although a degree of IR may be inherited, it may progress with additional factors such as obesity and a sedentary lifestyle. Hyperglycemia can worsen IR and insulin secretion because of effects such as decreased effective responsiveness of β -cells (possibly due to its effects on insulin gene expression) and decreased sensitization to insulin of target tissues. This can be due to multiple effects on cellular activity, including altered gene expression, enhanced oxidative stress, activation of serine/threonine kinases and protein kinase C, increased flux of glucose through the hexosamine biosynthesis pathway, and alterations to the polyol pathway.

T2DM and IGT appear to cluster in families, and twin studies have suggested that genetic makeup and certain ethnic backgrounds play important roles in an individual's susceptibility to T2DM (Tables 1–3, Figure 2).^{5,6,7,8,9}

Genome-wide association studies currently underway have led to the identification of many genetic loci and over 220 genes with associations that may confer an increased risk of developing T2DM.^{10,11} There appear to be associations between

fetal growth patterns, birth weight, paternal IR and/or T2DM, and umbilical cord insulin concentrations and later susceptibility to T2DM.^{12,13,14}

The convergence of factors such as physical inactivity, poor dietary habits, and obesity may increase the risk of developing diabetes in a genetically predisposed individual (Table 1, Fig. 2).¹⁵ Increased weight gain and decreased physical activity appear to affect insulin action and/or insulin secretion, particularly in populations that have rapidly adopted a Western lifestyle.¹⁶ Epidemiological studies are beginning to suggest that the levels of some compounds may lead to alterations of blood glucose in animal models.^{17,18}

CAUSES OF INSULIN RESISTANCE

IR may be defined as a subnormal biological response to a given concentration of insulin. The subnormal biological response could be due to the inability of plasma insulin to bind to its receptor or the presence of a postreceptor binding defect. IR can exist in many organs and tissues such as the liver and

Table 1. *Risk Predictors of Diabetes.*

Gender	Genetic defects
Female > Male	See Tables 2 and 3
Family history ^{5,6,7}	Environmental exposure
First-degree relative with T2DM (risk up to 40%)	Arsenic ¹⁸
Both parents with T2DM (risk up to 70%)	Bisphenol A ⁸⁷
Monozygotic twins > dizygotic twins	Tetrachlorodibenzo- <i>p</i> -dioxin ⁸⁸
	Phthalates ⁸⁹
Ethnicity (age-adjusted total prevalence) ^{91,92,93}	Obesity ^{65,80,91}
Caucasian: ~7%--8%	Duration of obesity ^{65,91}
African American: ~13%	Fat distribution ⁸⁶
Hispanic American: ~11%--12%	Waist/hip ratio ⁸⁶
American Indian: ~17%	Nutrition-related epigenetic changes
Medications	Endocrinopathies
Glucocorticoids	Diseases of the exocrine pancreas
Posttransplant immunosuppressives ⁹⁰	Cushing's syndrome
Calcineurin inhibitors ⁹⁰	Acromegaly

Abbreviation: T2DM, type 2 diabetes mellitus.

Table 2. *Genes Implicated in Diabetes.*

Peroxisome proliferator-activated receptor gamma (PPAR γ)
Peroxisome proliferator-activated receptor gamma coactivator 1 (PPARGC1)
Potassium inwardly-rectifying channel, subfamily J, member 11 (KCNJ11)
Transcription factor 7-like 2 (TCF7L2)
CDK5 regulatory subunit associated protein 1-like 1 (CDKAL1)
Hematopoietically expressed homeobox (HEX)
Solute carrier family 30, member 8 (SLC30A8) and solute carrier family 2, member 1 (SLC2A1)
Chromatin-remodeling protein 11 (Chr11)
Glycogen synthase 1 (GYS1)
Insulin receptor substrate 1 (IRS-1)
Insulin (INS)
KCNJ11
ATP-binding cassette, subfamily C, member 8 (ABCC8)
Calpain 10 (CAPN10)
Insulin-like growth factor 2 messenger RNA binding protein 2 (IGF2BP2)
Cyclin-dependent kinase inhibitor 2A/B (CDKN2A/B)
Fat mass and obesity associated (FTO)

NOTE: This table was adapted from *Lancet*⁹⁵, *Diabetes Report*⁹⁶, and *Diabetes Res Clin Pract.*⁹⁷

muscle. Some of the results of IR include the overproduction of glucose by the liver (despite fasting hyperinsulinemia and hyperglycemia) and decreased glucose clearance by peripheral tissues.^{19,20} In subjects with T2DM, modifications to the insulin receptor signal transduction pathway can occur and subsequently affect IR. This can affect the degree of translocation of the glucose transporter to the plasma membrane.

Table 3. *Insulin Secretory Defects in Type 2 Diabetes.*

Decreased insulin secretory capacity
Decreased β -cell sensitivity to glucose
Loss of first-phase insulin secretion to glucose
Increased release of proinsulin and its split products
Disruption of the normal pattern of pulsatile secretion
Morphological changes (decreased β -cell mass and amyloid deposits)

Adipose cells also appear to play an important role in the development of IR. Adipose cells can release inflammatory cytokines and plasma free fatty acids (FFAs). Elevated FFA levels appear to help stimulate gluconeogenesis, induce hepatic IR (through inhibition of the insulin signal transduction system), and decrease peripheral glucose clearance because of inhibitory effects on multiple steps in the metabolism of glucose and insulin action. The role of adipose tissue and proteins such as adiponectin and retinol-binding protein 4 in inducing IR are currently being elucidated.^{21,22,23,24,25}

Modifications to the Insulin Receptor Pathway

The decreased biological response to insulin could potentially be due to impairments in the insulin receptor signal transduction pathways. This can ultimately affect the glucose uptake by the cells and result in elevated plasma glucose. In the receptor signal transduction pathway, the activated insulin receptor phosphorylates intracellular substrates such as insulin receptor substrate (IRS) proteins through tyrosine phosphorylation. IRS proteins interact with phosphatidylinositol kinase (eg, phosphoinositide

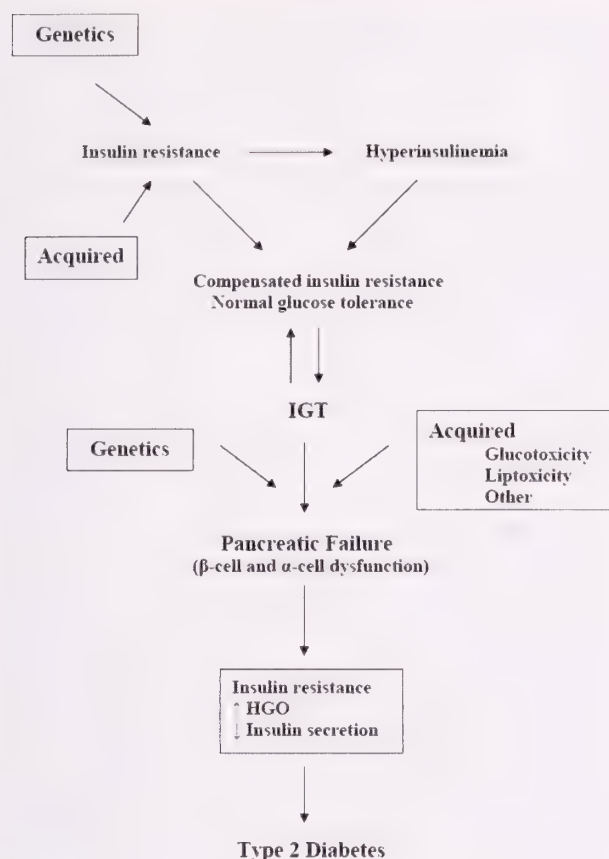


Fig 2. Proposed etiology for the development of type 2 diabetes mellitus.

3-kinase in muscle).²⁶ A downstream effect of this interaction is the activation of the Akt-2/PKB β pathway. Akt-2 (V-akt murine thymoma viral oncogene homolog 2; a member of the AKT protein family) is involved in signal transduction. Akt members are also called protein kinases B (PKB) and Akt-2 is analogous to PKB β . This activation step is important for the insulin-induced translocation of glucose transporter 4 (GLUT4) to the plasma membrane.²⁶ The development of IR may involve a common step or a combination of relatively minor defects in the described insulin-signaling pathway. Postreceptor defects could involve any of the insulin signal transduction pathway molecules.

A combination of relatively minor defects in the signal pathway appears to act synergistically to cause IR and glucose intolerance.²⁷ In T2DM, it has been noted that the ability of insulin to stimulate IRS-1 phosphorylation on tyrosines appears to be impaired in both adipocytes and skeletal muscle.²⁸ Elevations in glucose, FFAs, and inflammatory mediators [eg, tumor necrosis factor alpha (TNF- α), adipokines, and interleukin 6] also appear to indirectly affect

signal mediators. Elevations in protein kinase C, c-Jun N-terminal kinase, and I kappa β kinase appear to subsequently reduce downstream IRS-1 signaling and reduce GLUT4 transport.

Inflammation

Inflammation may also play a role in the pathogenesis of IR. Older approaches used in diagnosis (eg, increased acute-phase reactants: fibrinogen and C-reactive protein) and treatment (eg, salicylates) have long suggested a possible role of inflammation in T2DM. Adipose tissue may play an important role in the development of inflammation. The development and progression of IR may be related to substances released from adipocytes, such as adiponectin, and proinflammatory adipose tissue-derived cytokines (eg, resistin and TNF- α), as seen in obesity. TNF- α and plasminogen activating inhibitor-1 levels appear to be associated with IR and diabetes.^{29,30,31}

TNF- α enhances adipocyte lipolysis, which further increases nonesterified FFAs and subsequently increases IR by direct negative effects on insulin signaling pathways.^{32,33,34} FFAs also appear to serve as ligands for the toll-like receptors (eg, toll-like receptor 4), which appear to have a role in increased cytokine release.^{35,36} This may be another link connecting inflammatory pathways and IR. The abdominal fat-derived cytokines and lipids may enter the circulation through the portal system and contribute to hepatic inflammation and peripheral IR. The skeletal muscle does not appear to be a source of adipose tissue-related inflammatory cytokine activation. Skeletal muscle has been shown to develop IR from high levels of circulating fatty acids, which appear to disrupt insulin signaling pathways.³⁷

INSULIN SECRETION

β -Cells are capable of sensing plasma glucose levels and releasing insulin. Rising plasma glucose levels enters the β -cells through the GLUT2 transporter and subsequently lead to a rise in the β -cell adenosine triphosphate (ATP)/adenosine diphosphate (ADP) ratio. The rise in the ATP/ADP ratio initiates steps that lead to a depolarization of the cell membrane, an influx in calcium ions, and subsequent exocytotic release of insulin from storage granules. β -Cells attempt to maintain normoglycemia by modifying insulin secretion to affect the peripheral glucose disposal rates despite underlying IR. The ability to secrete adequate amounts of insulin depends on β -cell function, β -cell insulin receptor interactions, and mass³⁸. In NGT subjects, insulin secretion is often

increased in the presence of hyperglycemia in order to compensate for underlying IR.²

The insulin secretion ability appears variable in IGT and T2DM subjects. The insulin secretion ability is affected by the heterogeneity of the baseline insulin secretion capacity, glucotoxicity from chronic hyperglycemia, the confounding effects of obesity and circulating FFAs, and preexisting IR. Subjects with IGT, when compared with NGT subjects who are matched for degrees of IR and obesity, secrete less insulin at any given glucose level (Table 3).^{4,39} Islet β -cell function declines as subjects advance from IGT to T2DM^{40,41}. It further declines during established T2DM.^{42–45}

Both the amount of insulin secreted by β -cells and the phases of insulin secretion are affected in T2DM. NGT and IGT subjects, matched for body mass index and rendered with similar glucose levels by infusions of glucose, have higher circulating insulin levels than those with T2DM. Even though basal insulin levels appear to be elevated and maintained in the presence of fasting hyperglycemia in T2DM, total insulin levels may be confounded by the fact that proinsulin and its intermediates cross-react with insulin in most immunoassays. The cross-reaction of proinsulin and its intermediates often may cause the plasma concentration of true insulin to be overestimated. In T2DM, the progressive loss of the first-phase insulin response and delayed second-phase insulin response to meals may contribute to postprandial hyperglycemia with exaggerated insulin response 2 to 4 hours after a meal.⁴³ This progressive deterioration in β -cell function appears to be one of the main reasons that patients who are initially controlled on a single oral agent may later require additional therapy to achieve glycemic control.

β -Cell defects in T2DM are relatively specific for glucose. The insulin response to other stimuli, such as arginine, may be nearly normal in patients who have elevated fasting glucose levels. The ability of hyperglycemia to enhance the response to nonglucose stimuli is impaired in parallel with the decline in insulin secretion to glucose when patients and controls are studied at the same glucose level.⁴⁴ The relative preservation of insulin responses to certain amino acids and insulinogenic gut peptides (incretins) that are released during meal absorption may explain why the insulin response to mixed meals is better than the response to oral glucose.

The Incretin Effect

The β -cell secretion of insulin is greater after the oral administration of glucose than after the intravenous administration of glucose in NGT

subjects. This difference in insulin secretion is called the incretin effect.^{45,46} Normally, incretins are reduced in the fasting state and increase rapidly after a meal. The 2 major incretins in humans are glucagon-like peptide 1 (GLP-1; Figure 1) and glucose-dependent insulintropic peptide (GIP)⁴⁷. GIP (released by K-cells in the small intestine) and GLP-1 (mainly produced by L-cells in the distal intestine) are released during nutrient absorption in the gastrointestinal tract. They play an integral role in glucose homeostasis by augmenting insulin secretion. The incretin hormones GLP-1 and GIP are inactivated by the enzyme dipeptidyl peptidase 4⁴⁸. Although GLP-1 receptors appear more ubiquitous, GIP receptors are more often located in the pancreas and small intestine. The incretins are glucose-dependent and interact with G-protein-coupled receptors to activate molecular pathways.

In the presence of IGT and T2DM, the incretin effect is affected adversely. There appears to be a reduction in the nutrient-mediated secretion and variation in the level or action of GIP and GLP-1 after the ingestion of a mixed meal.^{49,50} In subjects with T2DM, there appears to be decreased responsiveness to the presence of GIP. Unlike GIP, GLP-1 is able to suppress glucagon secretion after a meal.^{51,52} GLP-1 also has other effects, such as growth factor-like effects on the β -cell stimulation of insulin gene expression and biosynthesis of insulin and restoration of glucose competence in glucose-resistant β -cells.^{53,54} Even though the level of GLP-1 decreases in T2DM, its insulintropic action is preserved. The preserved responsiveness to GLP-1 in subjects as they progress to T2DM has made GLP-1 a focus of antidiabetic therapy.⁵⁵ The decrease in the effective action of incretins in T2DM may help contribute to decreased β -cell mass and insulin secretion.

Possible Causes of β -Cell Dysfunction

The ability of β -cells to secrete insulin is dependent on their mass and function. NGT subjects who underwent partial pancreatectomy during the early 1990s as part of transplant programs subsequently developed IGT and T2DM. The β -cell mass may be reduced (by 20%–50%) in patients with longstanding T2DM because of increased apoptosis.⁴⁴ Fewer β -cells could be a result of longstanding diabetes. A decrease in β -cell numbers could lead to decreased function of the remaining β -cells. Studies suggest that in the absence of IR, more than 80% of β -cells must be lost before insulinopenic diabetes develops. There is some controversy about the degree to which β -cell mass is permanently decreased in light of *in vitro*

studies suggesting that the total β -cell mass may not be fixed at birth and that there may be some degree of β -cell regeneration.⁵⁶

β -cell insulin secretion capacity appears to be an important factor that declines in subjects as they progress from NGT to IGT and finally T2DM. Some of these defects may be hereditary, whereas others are acquired. Nutrient excess (glucose and FFAs) appears to negatively affect β -cell function through numerous mechanisms, including alterations in metabolic pathways, increases in intracellular calcium, effects on K-ATP channels, endoplasmic reticulum stress, and the generation of reactive oxygen species. The alleviation of glucotoxicity through improved long-term glycemic control may improve first-phase insulin response and defective glucose recognition by β -cells in some T2DM subjects.⁵⁷ Chronically increased uptake of FFAs by islet β -cells and/or a defect of β -cell FFA metabolism can also lead to islet lipid deposition, which contributes to the decline in insulin secretion in T2DM (Figure 1).^{58,59} The precise mechanisms underlying the lipotoxic effects during the development of T2DM are a matter of much debate.

Amyloid deposition in T2DM may also play a role in the reduction of β -cell section capacity.⁵⁹ Patients with T2DM have been noted to have increased islet amyloid [from islet amyloid polypeptide (IAPP)]. IAPP may increase because of increased production and/or decreased clearance. IAPP is stored in insulin secretory granules and is cosecreted with insulin in a 1:10 ratio.⁶⁰ It is normally released as a response to meals and appears to lower glucagon secretion and the rate of appearance of glucose after a meal.⁶¹ Subjects with T2DM have a lower first-phase serum IAPP response. It appears that increased levels of IAPP may accumulate in the β -cells and adversely affect them. IAPP-derived amyloid has the propensity to form membrane insoluble toxic oligomers that may be deposited between the capillaries, and cells, and within invaginations of the β -cell membrane. This may contribute to insulin secretory defects by acting as a barrier to the diffusion of nutrients and hormones, interfering with cell membrane function, and hence affecting glucose signaling and secretory granule release. There are studies suggesting that intracellular toxic oligomers, rather than extracellular deposits, may also play a role in the loss of β -cells in T2DM.⁶²

GLUCOSE METABOLISM IN TYPE 2 DIABETES MELLITUS

In normal glucose-tolerant subjects, the body tries to maintain normoglycemia in both the fasting and

postprandial states. In the fasting state, the plasma glucose levels are affected mainly by basal glucose levels and by the production of glucose by the liver. In the postprandial state, plasma glucose levels are affected by basal glucose levels and the breakdown of carbohydrates absorbed through digestion. The body attempts to maintain normoglycemia and helps transport glucose into cells through adequate insulin secretion. Glucose is mainly transported into cells by glucose transporters (eg, GLUT4 found in muscle and fat cells).

In T2DM, the body is not able to maintain normoglycemia. There is impaired glucose uptake due to the inadequate compensation of insulin secretion for IR. Plasma glucose levels may be increased not only because of the worsening compensation but also because of dysregulation of glucose production by the liver. The plasma glucose levels are also affected by ventromedial hypothalamic glucose-sensing neurons⁶³. The changes to ventromedial hypothalamic glucose-sensing neurons in T2DM can indirectly lead to increased dysregulation of gluconeogenesis. A decreased rate of muscle glycogen synthesis has been directly shown in T2DM. This may be partly due to impaired glycogen synthase and pyruvate dehydrogenase. The magnitude of this defect may correlate with an increased impairment of whole body glucose uptake in T2DM.^{33,37,64,65} Glucose transport through glucose transporters and enhanced non-insulin-mediated glucose uptake by tissues other than brain are unable to fully lower elevated plasma glucose levels.

Fasting State

Glucose can enter cells through insulin-dependent mechanisms and insulin-independent mechanisms. In a fasting state, insulin levels are low, and 75% to 85% of glucose uptake is non-insulin-mediated. Insulin-independent glucose uptake by the brain accounts for a large part of this basal uptake, whereas skeletal muscle accounts for only 15% to 20% of basal glucose uptake. This mass action glucose uptake mechanism is not affected in T2DM. The development of IR appears to have relatively little effect on basal glucose utilization rates. Fasting glucose levels can increase substantially if there is increased production of glucose that can enter the circulation primarily from the liver. In the setting of IR and impaired insulin secretion, insulin-mediated glucose uptake cannot increase appropriately in response to an increase in hepatic glucose output (HGO). The increased glucose entry into the circulation can be matched by increased glucose disposal in the presence of fasting hyperglycemia to a certain extent. An increase

in HGO likely contributes to the initial rise in plasma glucose and to fasting hyperglycemia in T2DM.

Hepatic IR begins to appear in IGT subjects and becomes more prevalent in T2DM. Although IGT and early T2DM subjects may have normal basal rates of HGO, in T2DM subjects with fasting plasma glucose > 140 mg/dL, the basal HGO tends to be modestly increased.³ In early IGT and T2DM, increased secretion of insulin initially helps restore normal levels of HGO by establishing a new steady state in which fasting glucose and insulin levels are elevated, but HGO is within the normal range, albeit not normal for the prevailing insulin and glucose levels. Studies suggest that gluconeogenesis is increased in T2DM patients who have fasting hyperglycemia.^{66,67} The presence of inappropriate glucagon secretion (due to reduced suppression of glucagon) appears to contribute to elevated plasma glucose levels in T2DM.

Recent studies have shown that the pathway of glycogen synthesis is also adversely affected. It appears that defects in the activation of glucokinase and dephosphorylation of glycogen phosphorylase may be contributing factors to the dysregulation of hepatic glucose metabolism in T2DM.^{68,69} Abnormal bile acid metabolism noted in T2DM patients may also be a factor in the increased HGO. In normal metabolism, bile acids have a role in decreasing gluconeogenesis, increasing glycogenesis and incretin release.⁷⁰ Although there is abnormal bile acid metabolism in T2DM, the mechanisms through which bile acids affect these pathways are still being elucidated.⁷¹

Pathogenesis of Postprandial Hyperglycemia

Several factors contribute to postprandial hyperglycemia in T2DM. First, the total amount of glucose entering the systemic circulation is increased because of impaired suppression of HGO. Second, the efficiency of glucose removal is reduced because of muscle IR and relative or absolute insulin deficiency. Third, hepatic uptake of glucose is impaired, although this defect makes a relatively small contribution.

The glucose homeostasis system has 2 sources of glucose input, HGO and glucose from the gastrointestinal tract, in the postprandial state, unlike in the fasting state, in which HGO is the primary source of glucose. The suppression of HGO after a meal is impaired, and subsequently, rising HGO contributes to postprandial glucose excursions.^{71,72,73,74}

This occurs because of the dysregulation of α -cells due to their refractoriness to the inhibitory effect of either hyperinsulinemia or chronic exposure to hyperglycemia, which may impair their glucose-sensing abilities. This leads to a decrease in the insulin-to-glucagon ratio in the postprandial state.^{75,76,77,78} A lack of glucagon suppression after carbohydrate ingestion causes postprandial hyperglycemia in healthy nondiabetic subjects. This likely contributes to postprandial hyperglycemia in T2DM subjects.^{79,80,81}

The lowering of endogenous glucose production through the administration of exenatide in T2DM human and animal studies seems to suggest that GLP-1 may have a role in not only increasing insulin secretion but also indirectly modulating glucagon secretion.⁸² There may exist a reciprocal interplay between glucagon, certain incretin hormones, and other intra-islet cell products via the islet microcirculation, which may also have a role in regulating glucagon secretion.^{83,84} Patients with IGT or T2DM have alterations in several of these factors, including glucose levels, insulin levels, and intra-islet cell messengers.

CONCLUSION

The development of T2DM appears to involve defects in insulin action and secretion. Genetic and environmental factors play an important role in the development of IR and defects in insulin secretion. Epidemiological and clinical evidence suggests that, in most populations, the interplay between IR and defects in insulin secretion initiates a sequence of events leading to T2DM. Glucose levels remain normal or only mildly impaired when islet β -cells remain able to compensate with sufficient insulin output to overcome IR. A progressive decline in β -cell compensation in a subset of insulin-resistant subjects eventually leads to overt hyperglycemia. IR affects the main insulin target tissues, namely, skeletal muscle, the liver, and adipose tissue. IR is likely due to the interplay of genetic factors, modifications to the insulin signaling pathway, and inflammation. Insulin secretion is affected by multiple factors, including the baseline secretion capacity, incretin effect, nutrient excess of glucose and FFAs, and amyloid deposition.

The interplay between IR and impaired insulin secretion leads to metabolic abnormalities in target tissues. Some of the metabolic abnormalities include decreased insulin-stimulated skeletal muscle glucose uptake, overproduction of glucose by the liver, and impaired suppression of adipocyte lipolysis. The

plasma glucose is elevated during fasting states in subjects with T2DM. This is partly due to the overproduction of glucose by the liver and glucagon dysregulation. Impaired insulin-stimulated muscle glucose uptake, dysregulation of glucagon secretion, and impaired suppression of gluconeogenesis are important causes of postprandial hyperglycemia. Elevated fasting and 24-hour circulating plasma FFA levels appear to also have an indirect role in elevated plasma glucose because of their effects on muscle IR and dysregulation of glucose production by the liver.

Current research is helping to clarify the complex processes that are associated with T2DM. An improved understanding of T2DM is needed to help clinicians prevent the development and progression of both microvascular and macrovascular complications associated with T2DM. Furthermore, a better understanding of T2DM will aid in the development of new treatments for T2DM.

ACKNOWLEDGMENTS

Diabetes research at Tulane University Health Sciences Center is supported in part by the Susan Harling Robinson Fellowship in Diabetes Research and the Tullis-Tulane Alumni Chair in Diabetes. Jennifer John-Kalarickal, Vivian A. Fonseca, and the Tulane Section of Endocrinology are also supported in part by the American Diabetes Association and the National Institutes of Health (the Action To Control Cardiovascular Risk in Diabetes Trial and the Targeting Inflammation Using Salsalate in Type 2 Diabetes Trial).

DISCLOSURES

Potential conflict of interest: Jennifer John-Kalarickal, Vivian A. Fonseca, and Tulane University have received research grants from Glaxo Smith Kline, Novartis, Takeda, Astra Zeneca, Pfizer, Sanofi-Aventis, Eli Lilly, Amylin, Daiichi-Sankyo, Biondi, and Mannkind. Vivian A. Fonseca has received honoraria for consulting and lectures from Daiichi Sankyo, Glaxo Smith Kline, Novartis, Takeda, Pfizer, Sanofi-Aventis, Eli Lilly, and Novo Nordisk.

REFERENCES

1. Eriksson J, Franssila-Kallunki A, Ekstrand A, et al. Early metabolic defects in persons at increased risk for non-insulin dependent diabetes mellitus. *The New England Journal of Medicine* 1989; 321: 337–343.
2. Cavaghan M, Ehrmann D, Polonsky K. Interactions between insulin resistance and insulin secretion in the development of glucose intolerance. *J Clin Invest* 2000; 106: 329–333.
3. Stern M. Strategies and prospects for finding insulin resistance genes. *J Clin Invest* 2000; 323–327.
4. Polonsky K, Sturis J, Bell G. Non-insulin-dependent diabetes mellitus: a genetically programmed failure of the beta cell to compensate for insulin resistance. *The New England Journal of Medicine* 1996; 334: 777–783.
5. Poulsen P., Ohm Kyvik K., Vaag A., Beck-Nielsen H. Heritability of Type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance—a population-based twin study. *Diabetologia* 1999; 42(2): 139–45.
6. Poulsen P., Levin K., Beck-Nielsen H., Vaag A. Age-dependent impact of zygosity and birth weight on insulin secretion and insulin action in twins. *Diabetologia* 2002; 45(12): 1649–57.
7. Groop LC, Tuomi T. Non-insulin-dependent diabetes mellitus—a collision between thrifty genes and an affluent society. *Ann Med* 1997; 29: 37–53.
8. Condon J, Shaw JE, Luciano M, et al. A study of diabetes mellitus within a large sample of Australian twins. *Twin Res Hum Genet* 2008; 11(1): 28–40.
9. Froguel P, Velho G. Genetic determinants of T2DM. *Recent Prog Horm Res* 2001; 56: 91–105.
10. Gaulton KJ, Willer CJ, et al. Comprehensive Association Study of T2DM and Related Quantitative Traits with 222 Candidate Genes. *Diabetes*. 2008 Aug 4.
11. Florez JC. The Genetics of T2DM: A realistic Appraisal Circa 2008. *J Clin Endocrinol Metab*. 2008 Sep 9.
12. Bergvall N, Cnattingius S. Familial (shared environmental and genetic) factors and the foetal origins of cardiovascular diseases and T2DM: a review of the literature. *J Intern Med*. 2008.
13. Davey Smith G, Sterne JA, Tynelius P, Rasmussen F. Birth characteristics of offspring and parental diabetes: evidence for the fetal insulin hypothesis. *J Epidemiol Community Health* 2004; 58: 126–8.
14. Lindsay RS, Dabelea D, Roumain J, et al. T2DM and low birth weight: the role of paternal inheritance in the association of low birth weight and diabetes. *Diabetes* 2000; 49: 445–9.
15. Diabetes Prevention Program Research Group. Reduction in the Incidence of T2DM with Lifestyle Intervention or Metformin. *N Engl J Med* 2002; 346(6): 393–403.
16. Clark D. Physical activity efficacy and effectiveness among older adults and minorities. *Diabetes Care* 1997; 20: 1176–1182.
17. Lee DH, Lee IK, Jin SH, et al. Association between serum concentrations of persistent organic pollutants and insulin resistance among nondiabetic adults: results from the National Health and Nutrition Examination Survey 1999–2002. *Diabetes Care* 2007; 30(3): 622–8.
18. Navas-Acien A, Silbergeld EK, Pastor-Barriuso R, Guallar E. Arsenic exposure and prevalence of T2DM in US adults. *JAMA* 2008; 300(7): 814–22.
19. Kelley D, Mintun M, Watkins S, et al. The effect of non-Insulin dependent diabetes mellitus and obesity on glucose transport and phosphorylation in skeletal muscle. *J Clin Invest* 1996; 97: 2705–2713.
20. Cline G, Petersen K, Krssak M, et al. Impaired glucose transport as a cause of decreased insulin-stimulated

- muscle glycogen synthesis in T2DM. *The New England Journal of Medicine* 1999; 341: 240–246.
21. Gavi, S, Stuart, LM, Kelly, P, et al. Retinol-binding protein 4 is associated with insulin resistance and body fat distribution in nonobese subjects without T2DM. *J Clin Endocrinol Metab* 2007; 92: 1886.
 22. Rajala MW, Scherer PE. Minireview: The Adipocyte—At the Crossroads of Energy Homeostasis, Inflammation, and Atherosclerosis. *Endocrinology* 2003; 144(9): 3765–3773.
 23. Goldstein BJ, Scalia R. Adiponectin: A Novel Adipokine Linking Adipocytes and Vascular Function. *J Clin Endocrinol Metab* 2004; 89(6): 2563–2568.
 24. Semple RK, Cochran EK, Soos MA, et al. Plasma adiponectin as a marker of insulin receptor dysfunction: clinical utility in severe insulin resistance. *Diabetes Care* 2008; 31(5): 977–9.
 25. Weyer C, Funahashi T, Tanaka S, et al. Hypoadiponectinemia in obesity and T2DM: close association with insulin resistance and hyperinsulinemia. *J Clin Endocrinol Metab* 2001; 86: 1930.
 26. Youngren JF. Regulation of insulin receptor function. *Cell Mol Life Sci* 2007; 64(7–8): 873–91.
 27. Kadowaki T. Insights into insulin resistance and T2DM from knockout mouse models. *J Clin Invest* 2000; 106: 459–465.
 28. Cusi K, Katsumi M, Osman A, et al. Insulin resistance differentially affects the PI-3-kinase- and MAP kinase-mediated signaling in human muscle. *J Clin Invest* 2000; 105: 311–320.
 29. Kanaya AM, Wassel Fyr C, Vittinghoff E, et al. Adipocytokines and incident diabetes mellitus in older adults: the independent effect of plasminogen activator inhibitor 1. *Arch Intern Med* 2006; 166: 350.
 30. Zinman, B, Hanley, AJ, Harris, SB, et al. Circulating tumor necrosis factor- α concentrations in a Native Canadian population with high rates of T2DM mellitus. *J Clin Endocrinol Metab* 1999; 84: 272.
 31. Uysal KT, Wiesbrock SM, Marino MW, Hotamisligil GS. Protection from obesity-induced insulin resistance in mice lacking TNF- α function. *Nature* 1997; 389: 610.
 32. Yki-Jarvinen H. Glucose toxicity. *Endocrine Rev* 1992; 13: 415–431.
 33. Shulman G. Cellular mechanisms of insulin resistance in humans. *Am J Cardiol* 1999; 83: 3J–10J.
 34. Bergman R. Non-esterified fatty acids and the liver: why is insulin secreted into the portal vein? *Diabetologia* 2000; 43: 946–952.
 35. Tsukumo DM, Carvalho-Filho MA, Carnevali JB, et al. Loss-of-function mutation in Toll-like receptor 4 prevents diet-induced obesity and insulin resistance. *Diabetes* 2007; 56(8): 1986.
 36. Reyna SM, Ghosh S, Tantiwong P, et al. Elevated toll-like receptor 4 expression and signaling in muscle from insulin-resistant subjects. *Diabetes* 2008; 57(10): 2595–602.
 37. Guilherme A, Virbasius JV, Puri V, Czech MP. Adipocyte dysfunctions linking obesity to insulin resistance and T2DM. *Nat Rev Mol Cell Biol* 2008; 9(5): 367–77.
 38. Okada T, Liew CW, Hu J, et al. Insulin receptors in beta-cells are critical for islet compensatory growth response to insulin resistance. *Proc Natl Acad Sci U S A* 2007; 104(21): 8977–82.
 39. Larsson H, Ahren B. Islet dysfunction in insulin resistance involves impaired insulin secretion and increased glucagon secretion in postmenopausal women with impaired glucose tolerance. *Diabetes Care* 2000; 23: 650–657.
 40. DeFronza R, Bonadonna R, Ferrannini E. Pathogenesis of NIDDM. A balanced overview. *Diabetes Care* 1992; 15: 318–368.
 41. Bruce D, Chisholm D, Storlien L, Kraegen E. Physiological importance of deficiency in early prandial insulin secretion in noninsulin-dependent diabetes. *Diabetes* 1998; 37: 736–744.
 42. Kahn S. The importance of the beta-cell in the pathogenesis of T2DM mellitus. *American Journal of Medicine* 2000; 108(Supplement 6a): 2S–8S.
 43. Reaven G, Chen Y, Hollenbeck C, et al. Plasma insulin, C-peptide, and proinsulin concentrations in obese and nonobese individuals with varying degrees of glucose tolerance. *Journal of Clinical Endocrinology Metabolism* 1993; 76: 44–48.
 44. Porte DJ, Kahn S. Beta-cell dysfunction and failure in T2DM: potential mechanisms. *Diabetes* 2001; 50(Supplement 1): S160–S163.
 45. Nauck M, Stockmann F, Ebert R, Creutzfeldt W. Reduced incretin effect in Type 2 (non-insulin-dependent) diabetes. *Diabetologia* 1986; 29(1): 46–52.
 46. Nauck MA, Homberger E, Siegel EG et al. Incretin effects of increasing glucose loads in man calculated from venous insulin and C-peptide responses. *J Clin Endocrinol Metab* 1986; 63(2): 492–498.
 47. Fehmann HC, Goke R, Goke B. Cell and molecular biology of the incretin hormones glucagon-like peptide-I and glucose-dependent insulin releasing polypeptide. *Endocrine Reviews* 1995; 16(3): 390–410.
 48. Baggio LL and Drucker DJ. Biology of incretins: GLP-1 and GIP. *Gastroenterology* 2007; 132(6): 2131–2157.
 49. Faerch K, Vaag A, Holst JJ, et al. Impaired fasting glycaemia vs impaired glucose tolerance: similar impairment of pancreatic alpha and beta cell function but differential roles of incretin hormones and insulin action. *Diabetologia* 2008; 51(5): 853–61.
 50. Toft-Nielsen MB, Damholt MB, Madsbad S et al. Determinants of the Impaired Secretion of Glucagon-Like Peptide-1 in Type 2 Diabetic Patients. *J Clin Endocrinol Metab* 2001; 86(8): 3717–3723.
 51. Drucker DJ. Enhancing Incretin Action for the Treatment of T2DM. *Diabetes Care* 2003; 26(10): 2929–2940.
 52. Drucker DJ. Glucagon-Like Peptide-1 and the Islet {beta}-Cell: Augmentation of Cell Proliferation and Inhibition of Apoptosis. *Endocrinology* 2003; 144(12): 5145–5148.
 53. Buteau J. GLP-1 receptor signaling: effects on pancreatic beta-cell proliferation and survival. *Diabetes Metab* 2008; 34(Suppl 2): S73–7.
 54. Gautier JF, Fetita S, Sobngwi E, Salaün-Martin C. Biological actions of the incretins GIP and GLP-1 and therapeutic perspectives in patients with T2DM. *Diabetes Metab* 2005; 31(3 Pt 1): 233–242.
 55. Henry R, Wallace P, Olefsky J. The effects of weight loss on the mechanisms of hyperglycemia in obese noninsulin-dependent diabetes mellitus. *Diabetes* 1986; 35: 990–998.
 56. A. Khalaileh, T. Gonen-Gross, J. Magenheimer, et al. Determinants of pancreatic b-cell regeneration.

- Diabetes, Obesity and Metabolism* 10(Suppl. 1): 2008; 128–135.
57. Tushuizen ME, Bunck MC, Pouwels PJ et al. Pancreatic Fat Content and β -Cell Function in Men With and Without T2DM. *Diabetes Care* 2007; 30(11): 2916–2921.
 58. McGarry J, Dobbins R. Fatty acids, lipotoxicity and insulin secretion. *Diabetologia* 1999; 42: 128–138.
 59. Cooper G, Day A, Willis A, et al. Amylin and the amylin gene: structure, function and relationship to islet amyloid and to diabetes mellitus. *Biochimica et Biophysica Acta* 1989; 1014(3): 247–258.
 60. Westermark P, Johnson KH, O'Brien TD, Betsholtz C. Islet amyloid polypeptide—a novel controversy in diabetes research. *Diabetologia* 1992; 35(4): 297–303.
 61. Haataja L, Gurlo T, Huang CJ, Butler PC. Islet amyloid in T2DM, and the toxic oligomer hypothesis. *Endocr Rev* 2008; 29(3): 303–16.
 62. Gastaldelli A, Baldi S, Pettiti M, et al. Influence of obesity and T2DM on gluconeogenesis and glucose output in humans: a quantitative study. *Diabetes* 2000; 49: 1367–1373.
 63. Sherwin RS. Bringing light to the dark side of insulin: a journey across the blood-brain barrier. *Diabetes* 2008; 57(9): 2259–68.
 64. Poitout V, Robertson RP. Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocr Rev* 2008; 29(3): 351–66.
 65. Narayan KMV, Boyle JP, Thompson TJ, et al. Effect of BMI on Lifetime Risk for Diabetes in the U.S. *Diabetes Care* 2007; 30(6): 1562–1566.
 66. Wajngot A, Chandramouli V, Schumann W, et al. Quantitative contributions of gluconeogenesis to glucose production during fasting in T2DM mellitus. *Metabolism* 2001; 50: 47–52.
 67. Agius L. Glucokinase and molecular aspects of liver glycogen metabolism. *Biochem J* 2008; 414(1): 1–18.
 68. Mitrakou A, Kelley D, Veneman T et al. Contribution of abnormal muscle and liver glucose metabolism to post-prandial hyperglycemia in NIDDM. *Diabetes* 1990; 39(11): 1381–1390.
 69. Abdul-Ghani M, DeFronzo RA. Fasting Hyperglycemia Impairs Glucose- But Not Insulin-Mediated Suppression of Glucagon Secretion. *J Clin Endocrinol Metab* 2007; 92(5): 1778–1784.
 70. Thomas C, Pellicciari R, Pruzanski M, et al. Targeting bile-acid signalling for metabolic diseases. *Nat Rev Drug Discov* 2008; 7(8): 678–93.
 71. Pehling G, Tessari P, Gerich J, et al. Abnormal meal carbohydrate disposition in insulin-dependent diabetes. Relative contributions of endogenous glucose production and initial splanchnic uptake and effect of intensive insulin therapy. *J Clin Invest* 1984; 74: 985–991.
 72. Consoli A. Role of liver in pathophysiology of NIDDM. *Diabetes Care* 1992; 15: 430–441.
 73. Unger R. Glucagon physiology and pathophysiology in the light of new advances. *Diabetologia* 1985; 28: 574–578.
 74. Reaven G, Chen Y, Godlay A, et al. Documentation of hyperglucagonemia throughout the day in nonobese and obese patients with noninsulin-dependent diabetes mellitus. *Journal of Clinical Endocrinology Metabolism* 1987; 64: 106–110.
 75. Muller W, Faloon G, Aguilar-Parada E, Unger R. Abnormal β -cell function in diabetes. Response to carbohydrate and protein ingestion. *The New England Journal of Medicine* 1970; 283: 109–115.
 76. Dimitriadis G, Pehling G, Gerich J. Abnormal glucose modulation of islet α - and β -cell responses to arginine in non-insulin-dependent diabetes mellitus. *Diabetes* 1985; 34: 541–547.
 77. Baron A, Schaeffer L, Shragg P, Kolterman O. Role of hyperglucagonemia in maintenance of increased rates of hepatic glucose output in type II diabetics. *Diabetes* 1987; 36: 274–283.
 78. Shah P, Basu A, Basu R, Rizza R. Impact of lack of suppression of glucagon on glucose tolerance in humans. *Am J Physiol* 1999; 277: E283–E290.
 79. Shah P, Vella A, Basu A, et al. Lack of suppression of glucagon contributes to post-prandial hyperglycemia in subjects with type 2 diabetes mellitus. *Journal of Clinical Endocrinology Metabolism* 2000; 85: 4053–4059.
 80. Ahren B, Larsson H. Impaired glucose tolerance (IGT) is associated with reduced insulin-induced suppression of glucagon concentrations. *Diabetologia* 2001; 44: 1998–2003.
 81. Bensal P, Wang Q. Insulin as a physiological modulator of glucagon secretion. *Am J Physiol Endocrinol Metab* 2008.
 82. Raju, B. and Cryer, P. E. Loss of the decrement in intraislet insulin plausibly explains loss of the glucagon response to hypoglycemia in insulin-deficient diabetes. *Diabetes* 2005; 54: 757–764.
 83. Gromada J, Franklin I, and Wollheim, C. B. α -Cells of the endocrine pancreas: 35 years of research but the enigma remains. *Endocr. Rev.* 2007; 28: 84–116.
 84. Peterson K, Dufour S, Befroy D, et al. Impaired mitochondrial activity in the insulin-resistant offspring of patients with T2DM. *NEJM* 2004; 350: 664.
 85. DeFronzo RA, Ferrannini E. Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. *Diabetes Care* 1991; 14(3): 173–94.
 86. Chan JM, Rimm EB, Colditz GA, et al. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care* 1994; 17(9): 961–9.
 87. Lang IA, Galloway TS, Scarlett A, et al. Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults. *JAMA* 2008; 300(11): 1303–10.
 88. Carpenter DO. Environmental contaminants as risk factors for developing diabetes. *Rev Environ Health* 2008; 23: 59–74.
 89. Stahlhut RW, van Wijngaarden E, Dye TD, et al. Concentrations of urinary phthalate metabolites are associated with increased waist circumference and insulin resistance in adult U.S. males. *Environ Health Perspect* 2007; 115(6): 876–82.
 90. Penforis A, Kury-Paulin S. Immunosuppressive drug-induced diabetes. *Diabetes Metab* 2006; 32(5 Pt 2): 539–46.
 91. National Institute of Diabetes and Digestive and Kidney Diseases. *National Diabetes Statistics, 2007 fact sheet*. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health; 2008.
 92. Prevalence of diabetes among Hispanics—selected areas, 1998–2002. *Morb Mortal Wkly Rep* 2004; 53(40): 941–4.

93. Bennett, PH. Epidemiology of diabetes mellitus. In: Ellenberg and Rifkin's Diabetes Mellitus, Rifkin, H, Porte, D Jr (Eds.) New York: *Elsevier*; 1990; 363–377.
94. Carpenter DO. Environmental contaminants as risk factors for developing diabetes. *Rev Environ Health* 2008; 23: 59–74.
95. Stumvoll M, Goldstein BJ, van Haefen TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 2005; 365(9467): 1333–46.
96. Narayan. Diabetes Rep 2007.
97. Malecki MT. Genetics of type 2 diabetes mellitus. *Diabetes Res Clin Pract* 2005; 68(Suppl 1): S10–21.

Glycemic Targets for Patients with Type 2 Diabetes Mellitus

Ole-Petter R. Hamnvik, MB, BCh,^{1,2} and Graham T. McMahon, MD, MMSc^{2,3}

¹ Department of Medicine, Brigham and Women's Hospital, Boston, MA

² Department of Medicine, Harvard Medical School, Boston, MA

³ Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Boston, MA

ABSTRACT

Cardiovascular disease is the predominant cause of death in diabetic patients, and reducing the risk of cardiovascular disease in diabetics has recently been the focus of several highly publicized large trials, including ACCORD (Action To Control Cardiovascular Risk in Diabetes), ADVANCE (Action in Diabetes and Vascular Disease: Preterax and Diamicon MR Controlled Evaluation), and VADT (Veterans Affairs Diabetes Trial). These studies randomized high-risk diabetic patients into either intensive treatment or standard treatment. The glycemic control arm of ACCORD was terminated 17 months before the planned end of the study because of a finding of significantly increased all-cause and cardiovascular mortality in the intensive treatment group. These findings were not duplicated in either ADVANCE or VADT. Multiple possible explanations have been brought forward, including a higher incidence of death from unrecognized hypoglycemia, effects due to increased exposure to particular antidiabetic medications, adverse effects of rapid correction of hyperglycemia, weight gain, and differences in baseline characteristics. None of these were validated in post hoc analyses of the trial data, and the cause of the increased mortality remains elusive. Subgroup analyses suggest that those who start off with better control of their diabetes or without preexisting cardiovascular disease may have the most to gain from tight glycemic control. Reducing the risk of macrovascular disease and death in diabetic patients requires

not only attention to glucose control but also meticulous attention to control of nonglycemic risk factors, including hypertension, hyperlipidemia, smoking, lack of exercise, and unhealthy diet as well as timely prescription of medications with proven preventative benefits, such as aspirin and statins. *Mt Sinai J Med* 76:227–233, 2009. © 2009 Mount Sinai School of Medicine

Key Words: cardiovascular disease, diabetes mellitus, glycemic control.

Hyperglycemia increases the risk of both microvascular and macrovascular disease in patients with diabetes. Observational data from the United Kingdom Prospective Diabetes Study trial¹ and other studies² have demonstrated that the higher the hemoglobin A1c (HbA1c) level is, the higher the risk is of myocardial infarction and other cardiovascular events. Each 1% increase in HbA1c was associated with a 13% increase in the risk of coronary artery disease in a meta-analysis of 5 studies.² This association remains valid even when HbA1c values are within what is considered to be the normal range.³

The importance of glycemic control in preventing the microvascular complications of retinopathy, nephropathy, and neuropathy in patients with type 2 diabetes has been consistently demonstrated over the past 20 years.⁴ However, the benefit of glycemic control with respect to the macrovascular complications of diabetes is more controversial. Interventional studies designed to prove a cardiovascular benefit from lowering HbA1c have reached divergent conclusions. The Veterans Affairs Feasibility Trial showed a non-significant increase in the risk of cardiovascular events in the intensive-treatment group.⁵ The United Kingdom Prospective Diabetes Study trial, on the other hand, showed a decrease of borderline significance ($P = 0.052$) in the number of

Address Correspondence to:

Graham T. McMahon

Division of Endocrinology,
Diabetes, and Hypertension
Brigham and Women's Hospital
Boston, MA

Email: gmcmahon@partners.org

myocardial infarctions in the intensive-treatment group.¹¹

GLYCEMIC CONTROL AND CARDIOVASCULAR OUTCOMES IN THREE RECENT STUDIES

ACCORD (Action To Control Cardiovascular Risk in Diabetes)⁷ was initiated in North America in 2001. The aim was to determine strategies to reduce cardiovascular mortality in type 2 diabetics. The study's 10,251 participants were randomized into 3 substudies examining the utility of blood pressure control, treatment of hyperlipidemia, and treatment of hyperglycemia. The inclusion criteria for the study selected participants with a relatively high risk of cardiovascular events.⁸ Within the hyperglycemia arm, the goal HbA1c of those receiving intensive treatment was no more than 6.0%, although at the time of termination, the average HbA1c was 6.4% (from a baseline of 8.2%).

ADVANCE (Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation)⁹ also began enrolling patients in 2001, and it also aimed to investigate the effect of intensive glucose control and intensive blood pressure control on cardiovascular morbidity and mortality in type 2 diabetics. The patient population consisted of 11,140 participants. Like the ACCORD patients, these patients had a high risk of cardiovascular events. Like ACCORD, this trial also tested the effect of another intervention on cardiovascular events, in this case the use of a fixed combination of indapamide and perindopril for blood pressure control. The glycemic control arm of the study included an intensive control group based on the use of gliclazide (modified release) with other medications added as needed to achieve an HbA1c level of less than 6.5%.¹⁰

VADT (Veterans Affairs Diabetes Trial)¹¹ was a trial of glycemic control only that recruited high-risk patients with poorly controlled diabetes on oral antidiabetic monotherapy. All patients had their nonglycemic cardiovascular risk factors optimally treated. The goal HbA1c in the intensive control group was a normal value (a value of 6.9% was achieved), whereas the standard group was treated for a goal of 8.0% to 9.0% (Table 1).

CARDIOVASCULAR DEATHS AND INTENSIVE TREATMENT OF DIABETES

These 3 studies all aimed to establish whether tight glycemic control with goal HbA1c levels below what is currently recommended would reduce cardiovascular events in high-risk diabetic patients. Although they were all large, similarly designed, prospective randomized controlled trials with blind adjudication of events, they reached different conclusions.

In February 2008, the ACCORD trial investigators announced higher all-cause mortality in the intensive glycemic control group,¹² with 257 deaths in this group versus 203 deaths in the standard treatment group. On the advice of the independent data safety and monitoring board, the intensive glucose control arm was discontinued 17 months prematurely, although the blood pressure control and lipid control arms continued.⁷ In response to the cessation of the glycemia substudy in ACCORD, the investigators of the ADVANCE study announced and subsequently published the results from their trial.^{9,13}

In the ADVANCE study, which was published in the same issue of the *New England Journal of Medicine*, no increased mortality was noted in the intensive glycemic control group, with 498 deaths in the intensive treatment group versus 533 in the control group. The VADT study also found similar rates of all-cause mortality in the 2 groups (95 deaths in the standard-therapy group versus 102 in the intensive-therapy group, $P = 0.62$).¹¹

The finding of increased risk of death from any cause in the ACCORD trial seems to have been driven mainly by an increase in the number of deaths from cardiovascular causes (135 versus 94 deaths). This heterogeneous group consisted of deaths from myocardial infarction, congestive heart failure, procedures, arrhythmias, stroke, and other cardiovascular causes. In addition, there were 86 deaths (versus 67 in the control group) from "unexpected or presumed cardiovascular disease." The latter group has sparked much discussion: are these truly cardiovascular deaths, or could these be deaths caused by hypoglycemia or other noncardiovascular causes?

The rate of hypoglycemic events requiring medical assistance in the ACCORD trial was markedly higher in the tight control group compared to the standard control group: 538 events in the intensive therapy group versus 179 events in the standard therapy group. According to the post hoc analyses

Table 1. Comparison of the ACCORD, ADVANCE, and VADT Studies.

	Trial		
	ACCORD ⁷	ADVANCE ⁹	VADT ^{11,30,31}
Characteristics at enrollment			
Number of participants	10,251	11,140	1,791
Geographical setting	United States and Canada	Worldwide	United States
Mean glycated hemoglobin (%)	8.3	7.5	9.4
Mean age (years)	62	66	60
Females (%)	38	42	3
Mean duration of diabetes (years)	10 (median)	8	11.5
History of cardiovascular events (%)	35	32	40
Study interventions			
Target glycated hemoglobin (%)	<6.0 versus 7.0–7.9	<6.5 versus local guidelines	Normal versus 8.0–9.0
Duration of follow-up (years)	3.5	5	5.6
Medications prescribed (intensive group versus standard group)			
Metformin (%)	95 versus 87	74 versus 67	60 versus 55
Glimepiride (%)	78 versus 68	92 versus 59 (includes all sulfonylureas)	55 versus 45
Thiazolidinedione (%)	92 versus 58	17 versus 11	72 versus 62 (rosiglitazone only)
Insulin (%)	77 versus 55	41 versus 24	90 versus 74
Statin (%)	88 versus 88	46 versus 48	86 versus 83
Aspirin (%)	76 versus 76	57 versus 55	94 versus 91
Control of other cardiovascular risk factors (intensive group versus standard group)			
Low-density lipoprotein (mg/dL)	90.8 versus 90.6	102 versus 102	80 (both groups combined)
High-density lipoprotein (mg/dL)	N/A	48 versus 48	40 versus 41
Triglycerides (mg/dL)	N/A	128 versus 141	151 versus 159
Mean blood pressure (mm Hg)	126/67 versus 127/68	136/74 versus 138/78	127/68 versus 125/69
Smoking (%)	9.8 versus 9.9	8.3 versus 7.8	6.0 versus 9.7
Outcomes (intensive group versus standard group)			
Mean glycated hemoglobin at end of study (%)	6.4 versus 7.5*	6.5 versus 7.3*	6.9 versus 8.4*
Death from any cause (%)	5.0 versus 4.0*	8.9 versus 9.6	11.4 versus 10.6
Cardiovascular deaths (%)	2.6 versus 1.8*	4.5 versus 5.2	4.3 versus 3.2
Nonfatal myocardial infarction (%)	3.6 versus 4.6*	2.7 versus 2.8	7.2 versus 8.7
Nonfatal stroke (%)	1.3 versus 1.2	3.8 versus 3.8	3.1 versus 4.0
Severe hypoglycemia requiring medical assistance or with impaired level of consciousness (%/year)	3.1 versus 1.0*	0.7 versus 0.4*	12 versus 4*
Change in weight (kg)	+3.5 versus +0.4*	−0.1 versus −1.0*	+8.2 versus +4.1*

NOTE: To convert cholesterol values from milligrams per deciliter to millimoles per liter, multiply by 0.02586. To convert triglyceride values from milligrams per deciliter to millimoles per liter, multiply by 0.01129.

Abbreviations: ACCORD, Action To Control Cardiovascular Risk in Diabetes; ADVANCE, Action in Diabetes and Vascular Disease: Preterax and Diamicon MR Controlled Evaluation; N/A, not applicable; VADT, Veterans Affairs Diabetes Trial.

* In these outcomes, the difference between the intensive treatment group and the standard treatment group achieved statistical significance.

performed, people who experienced hypoglycemic events were not more likely to die than those who did not. However, this does not rule out hypoglycemia as a potential culprit for the increased mortality in this group because hypoglycemic death is difficult to diagnose at autopsy and may present

as sudden death. This is supported by the finding of increased sudden death in the intensive treatment group of VADT (11 versus 4 deaths, $P = 0.08$).¹¹ The etiology of hypoglycemic death is presumed to be overactivation of sympathetic nerves with resulting cardiac dysrhythmias,¹⁴ which may be

particularly important in people with established cardiovascular disease. Notably, there were fewer myocardial infarctions in the intensive therapy group of the ACCORD trial (186 versus 235 in the standard therapy group). Despite this, the cardiovascular mortality was higher. It is possible that hypoglycemia with its sympathetic surge triggered more arrhythmias in the intensive therapy group, and this would explain the apparent paradox of fewer cardiovascular events but higher mortality in those that did have a cardiovascular event.

In contrast, both ADVANCE and VADT had higher hypoglycemia rates in the intensive treatment group but no associated increase in mortality. Whether this suggests that there was no increased mortality due to hypoglycemia or that the adverse effect of hypoglycemia in these trials was partially balanced by some beneficial effect of glycemic control remains to be determined. The combination of sulfonylureas with insulin is known to cause higher rates of hypoglycemia and as such may be linked with higher mortality. Strategies to control hyperglycemia with less risk of hypoglycemia should be the subject of future trials.¹⁵

Differential use of medications between the intensive and standard therapy groups may also play a role in the differential in mortality. Although the intensive therapy group was exposed to higher doses of oral antihyperglycemics and insulin, post hoc analyses did not reveal any association with a particular medication or medication combination. In particular, there was no apparent relation to exposure to thiazolidinediones. Of interest, fewer patients in ADVANCE were exposed to insulin (41%) and thiazolidinediones (17%) than in ACCORD (77% and 92%, respectively). The increased exposure to thiazolidinediones, rosiglitazone in particular, may have adverse cardiac results as suggested by a recent meta-analysis of the cardiac effects of rosiglitazone,¹⁶ although this finding was not statistically significant in further analyses.¹⁷ The raised insulin levels caused by exogenous administration or by oral insulin secretagogues have been linked to arterial and endothelial dysfunction through a mechanism independent of blood glucose levels.¹⁸ However, the participants in VADT also had high exposure to insulin and thiazolidinediones—90% and 72%, respectively—with no evidence of increased mortality; this suggests that exposure to these medications is not sufficient to explain the differences between ACCORD and ADVANCE.

The velocity of correction of hyperglycemia has also been postulated to play a role. Although in ACCORD the goal HbA1c was achieved after 3 months, this took over 6 months in ADVANCE and

VADT. Rapid correction of hyperglycemia is known to worsen diabetic retinopathy,¹⁹ but it may also have unknown adverse vascular effects that would explain the findings in ACCORD.

Weight gain is common in diabetics who take insulin or secretagogues. In ACCORD, there was a significant difference in the amount of weight gained between the intensive and standard groups: 27.8% gained more than 10 kg in the intensive therapy group versus 14.1% in the standard therapy group. Also, although the weight gain averaged 3.5 kg in ACCORD and 8.2 kg in VADT, there was a tendency toward weight loss in ADVANCE (although the average weight in the intensive treatment group was still 0.7 kg higher than that of the standard group). However, preliminary exploratory analyses did not find a link with increased mortality in those who gained more weight, although previous population-based analyses have shown increased mortality in patients with diabetes with an elevated body mass index.²⁰ Weight gain is usually more marked in patients receiving insulin or a thiazolidinedione and may contribute to any differential effect of a medication class that may be found. Also, thiazolidinediones may cause weight gain from fluid retention, which could increase the risk of heart failure and thereby increase mortality.

Other proposed explanations for the surprising findings in ACCORD include differing clinical characteristics at baseline. Post hoc analysis looking at differences in baseline characteristics, risk factors for hypoglycemic events, cointerventions from the blood pressure or lipid control arms, achieved blood pressure, achieved HbA1c, gender, or previous cardiovascular events did not demonstrate a consistent signal that might indicate high risk from 1 specific group.

Nevertheless, the different findings in the 3 trials have been attributed to the variance within their subjects. HbA1c values at enrollment differed: 8.3% in ACCORD, 7.5% in ADVANCE, and 9.4% in VADT. In an attempt to correct for this, the ADVANCE investigators performed a subgroup analysis, looking at the patient group that had HbA1c values closer to those of the ACCORD participants. They did not find a difference in mortality between the 2 treatment strategies in this subgroup.

The finding of increased cardiovascular mortality in ACCORD remains unexplained, despite a multitude of theories.

BENEFITS OF GLYCEMIC CONTROL

The information about cardiovascular risk reduction from these trials has been eclipsed by the controversy

surrounding the mortality findings. However, the goal of these studies was to examine whether there was any cardiovascular benefit from tight glycemic control, and cardiovascular mortality was only 1 contributor to the composite primary endpoint. None of the 3 major trials showed a cardiovascular benefit of tight glycemic control. In ACCORD, the primary endpoint (consisting of nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular causes) occurred in 352 participants in the tight control group versus 371 patients in the standard group ($P = 0.16$). In VADT, the primary endpoint (similar to that of ACCORD but with the additional inclusion of congestive heart failure and surgery/amputation for peripheral vascular disease) occurred in 235 patients in the intensive treatment group versus 264 in the standard-therapy group ($P = 0.14$). In ADVANCE, the primary endpoint was unusually broad and included microvascular events (such as worsening nephropathy or retinopathy). ADVANCE showed a statistically significant reduction in the occurrence of this composite endpoint in the intensive control group (1009 patients in the intensive group versus 1116 in the standard treatment group). This was largely driven by the reduction in microvascular events (particularly nephropathy), and this trial therefore confirmed the findings of previous trials. However, the data for cardiovascular events from ACCORD show no benefit of glycemic control: 557 participants in the intensive control group had a major macrovascular event versus 590 in the standard group (this was not significant). The reduction in nephropathy in ADVANCE is noteworthy, as nephropathy and microalbuminuria have been shown to be long-term risk factors for cardiovascular disease.²¹

In ACCORD, a nonsignificant reduction in composite primary outcome was noted after 3 years. In ADVANCE, a divergence appeared between the 2 groups in the incidence of major macrovascular events at 5 years and in the incidence of death at 4.5 years, in both cases favoring the intensive therapy group. These findings are not significant, but the studies were not adequately powered to detect a late favorable effect of intensive glucose lowering. Follow-up of the cohorts after the end of the trials may be informative because many other trials have shown that cardiovascular protection accrues in the long term.^{22,25}

The patients studied in these trials may not be the patients that have the most to gain from tight glycemic control. In ACCORD, subgroup analyses noted that there were fewer fatal and nonfatal cardiovascular events in those with no

Table 2. *Key Points.*

- Control of hyperglycemia in diabetes has been proven to reduce microvascular complications, but not cardiovascular outcomes.
- The ACCORD trial showed increased mortality in patients with tight glycemic control, a finding not replicated in the ADVANCE and VADT trials.
- Unrecognized hypoglycemic episodes are one of many possible explanations for the increased mortality in ACCORD.
- Follow-up studies and subgroup analyses may identify patients that are more likely to benefit from tighter glycemic control.
- The mainstay of optimizing cardiovascular outcomes in diabetes remains comprehensive focus on modifiable risk factors.

history of previous cardiovascular events and in those with a baseline HbA1c of less than 8%. This suggests that patients may potentially derive cardiovascular benefit from tight glycemic therapy early on. In addition, the information from other ongoing trials, such as BARI 2D (Bypass Angioplasty Revascularization Investigation 2 Diabetes),²⁴ ORIGIN (Outcome Reduction with an Initial Glargine Intervention),²⁵ and HEART2D (Hyperglycemia and Its Effect After Acute Myocardial Infarction on Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus),²⁶ will help clarify these findings.

CONCLUSION

Each of the 3 recently publicized trials of glycemic control in patients with type 2 diabetes is proof of the successes of modern diabetes management. In comparison with older trials, the incidence of cardiovascular events in a high-risk population was much lower than expected. Although the event rate for macrovascular events in these studies averaged around 20 per 1000 patient years, this rate has been as high as 50 per 1000 patient years in other epidemiological studies.²⁷ This is largely a result of the current focus on managing all cardiovascular risk factors in patients with type 2 diabetes. This multifaceted comprehensive approach includes control of smoking, hypertension, hyperlipidemia, lack of exercise, and poor diet. The integrated approach to cardiovascular risk reduction has demonstrated efficacy in studies such as the Steno-2.²⁸ Individual components (including blood pressure control in the blood pressure arm of the ADVANCE trial²⁹) have repeatedly been shown to have benefit. It is notable that of all of the goals, glycemic control

appears to be the most challenging to achieve: in Steno-2, few patients achieved goal HbA1c levels, but most patients achieved blood pressure and lipid targets. The VADT trial was unique among the 3 trials in its focus on achieving blood pressure and lipid goals and managing other lifestyle factors.

Aspirin should be indicated for most of the participants in these trials because of the high prevalence of other cardiac risk factors. Despite the widespread acknowledgment of the efficacy of aspirin in both secondary and primary prevention, only 75% of the participants in ACCORD and only 55% of the ADVANCE participants were taking aspirin. The numbers for statins were 56% and 87%, respectively. VADT did better, with numbers of 94% and 86% for aspirin and statins, respectively. The low use of these medications suggests that the risk of cardiovascular death can be further lowered by the generation of systems of care to ensure that eligible patients receive these effective therapies. The trials prove that continued effort is still required in order to promote adherence to published guidelines for cardiovascular risk factor modification.

For now, clinicians should target a glycated hemoglobin value of no more than 7.0%. This is a goal that has been proven to be beneficial in the prevention of renal and retinal disease, although there is no good evidence that targeting this measure will reduce macrovascular disease, at least in the first few years. Hypoglycemia should be carefully minimized, particularly in patients with established macrovascular disease. Smoking cessation, dietary and exercise counseling, blood pressure control, aspirin, and lipid control with statins should remain other areas of continual focus in order to reduce the risk of cardiovascular events in our patients with diabetes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Stratton IM, Adler AI, Neil HA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ* 2000; 321: 405–412.
2. Selvin E, Marinopoulos S, Berkenblit G, et al. Meta-analysis: glycosylated hemoglobin and cardiovascular disease in diabetes mellitus. *Ann Intern Med* 2004; 141: 421–431.
3. Khaw KT, Wareham N, Bingham S, et al. Association of hemoglobin A1c with cardiovascular disease and mortality in adults: the European Prospective Investigation into Cancer in Norfolk. *Ann Intern Med* 2004; 141: 413–420.
4. Ohkubo Y, Kishikawa H, Araki E, et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. *Diabetes Res Clin Pract* 1995; 28: 103–117.
5. Abraira C, Colwell J, Nuttall F, et al. Cardiovascular events and correlates in the Veterans Affairs Diabetes Feasibility Trial. Veterans Affairs Cooperative Study on Glycemic Control and Complications in Type II Diabetes. *Arch Intern Med* 1997; 157: 181–188.
6. UK Prospective Diabetes Study Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; 352: 837–853.
7. Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 2008; 358: 2545–2559.
8. Goff DC Jr, Gerstein HC, Ginsberg HN, et al. Prevention of cardiovascular disease in persons with type 2 diabetes mellitus: current knowledge and rationale for the Action To Control Cardiovascular Risk in Diabetes (ACCORD) trial. *Am J Cardiol* 2007; 99: 4i–20i.
9. Patel A, MacMahon S, Chalmers J, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med* 2008; 358: 2560–2572.
10. Study rationale and design of ADVANCE: Action in Diabetes and Vascular Disease—Preterax and Diamicron MR Controlled Evaluation. *Diabetologia* 2001; 44: 1118–1120.
11. Duckworth W, Abraira C, Moritz T, et al. Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med* 2009; 360: 129–139.
12. For safety, NHLBI changes intensive blood sugar treatment strategy in clinical trial of diabetes and cardiovascular disease. Available at: <http://public.nhlbi.nih.gov/newsroom/home/getpressrelease.aspx?id=2551>. Accessed April 2009.
13. Major international diabetes study does not confirm increased risk of death reported by US trial. Available at: <http://www.advance-trial.com/static/html/virtual/contents.asp?p=39>. Accessed April 2009.
14. Lindstrom T, Jorfeldt L, Tegler L, Arnqvist HJ. Hypoglycaemia and cardiac arrhythmias in patients with type 2 diabetes mellitus. *Diabet Med* 1992; 9: 536–541.
15. Dluhy RG, McMahon GT. Intensive glycemic control in the ACCORD and ADVANCE trials. *N Engl J Med* 2008; 358: 2630–2633.
16. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; 356: 2457–2471.
17. Monami M, Marchionni N, Mannucci E. Winners and losers at the rosiglitazone gamble: a meta-analytical approach at the definition of the cardiovascular risk profile of rosiglitazone. *Diabetes Res Clin Pract* 2008; 82: 48–57.
18. Arcaro G, Cretti A, Balzano S, et al. Insulin causes endothelial dysfunction in humans: sites and mechanisms. *Circulation* 2002; 105: 576–582.

19. Early worsening of diabetic retinopathy in the Diabetes Control and Complications Trial. *Arch Ophthalmol* 1998; 116: 874–880.
20. Khalangot M, Tronko M, Kravchenko V, et al. Body mass index and the risk of total and cardiovascular mortality among patients with type 2 diabetes: a large prospective study in Ukraine. *Heart* 2009; 95: 454–460.
21. Gerstein HC, Mann JF, Yi Q, et al. Albuminuria and risk of cardiovascular events, death, and heart failure in diabetic and nondiabetic individuals. *JAMA* 2001; 286: 421–426.
22. Holman RR, Paul SK, Bethel MA, et al. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med* 2008; 359: 1577–1589.
23. Nathan DM, Cleary PA, Backlund JY, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med* 2005; 353: 2643–2653.
24. Magee MF, Isley WL. Rationale, design, and methods for glycemic control in the Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) trial. *Am J Cardiol* 2006; 97: 20G–30G.
25. Gerstein H, Yusuf S, Riddle MC, et al. Rationale, design, and baseline characteristics for a large international trial of cardiovascular disease prevention in people with dysglycemia: the ORIGIN trial (Outcome Reduction with an Initial Glargine Intervention). *Am Heart J* 2008; 155: 26–32.
26. Milicevic Z, Raz I, Strojek K, et al. Hyperglycemia and Its Effect After Acute Myocardial Infarction on Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus (HEART2D): study design. *J Diabetes Complications* 2005; 19: 80–87.
27. Booth GL, Kapral MK, Fung K, Tu JV. Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study. *Lancet* 2006; 368: 29–36.
28. Gaede P, Lund-Andersen H, Parving HH, Pedersen O. Effect of a multifactorial intervention on mortality in type 2 diabetes. *N Engl J Med* 2008; 358: 580–591.
29. Patel A, MacMahon S, Chalmers J, et al. Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the ADVANCE trial): a randomised controlled trial. *Lancet* 2007; 370: 829–840.
30. American Diabetes Association. Intense blood glucose control yields no significant effect on CVD reduction in VA Diabetes Trial. Available at: <http://professional.diabetes.org/UserFiles/File/Scientific%20Sessions/2008/Media%20Page/On%20Site%20Releases/VA%20Study%20FINAL2%20press%20release.pdf>. Accessed April 2009.
31. Abraira C, Duckworth WC, Moritz T. Glycaemic separation and risk factor control in the Veterans Affairs Diabetes Trial: an interim report. *Diabetes Obes Metab* 2009; 11: 150–156.

Balancing Risk and Benefit with Oral Hypoglycemic Drugs

Ole-Petter R. Hamnvik, MB, BCh,^{1,2} and Graham T. McMahon, MD, MMSc^{2,3}

¹ Department of Medicine, Brigham and Women's Hospital, Boston, MA

² Department of Medicine, Harvard Medical School, Boston, MA

³ Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Boston, MA

ABSTRACT

Clinicians are faced with an expansive array of treatment choices when caring for patients with type 2 diabetes. Because patient compliance may be affected when media sensationalism about controversial findings is misunderstood, we sought to clarify the recent controversy surrounding the cardiovascular and bone-health risks of thiazolidinediones, the risk of lactic acidosis with metformin, and the risk of hypoglycemia with oral therapies. The side effect profile of thiazolidinediones includes fluid retention, heart failure; and an increased risk of fracture. A recent controversial meta-analysis suggested that rosiglitazone increases the risk of myocardial infarction, which is possibly related to thiazolidinedione-induced lipid changes, weight gain, congestive heart failure, and anemia. Metformin is restricted to patients with normal renal function because of concerns that metformin may cause lactic acidosis. However, few cases of metformin-associated lactic acidosis have been reported, and most have occurred in patients with other reasons for developing lactic acidosis, such as sepsis or renal failure. Although the use of metformin continues to increase, observational studies have not been able to demonstrate an increased incidence of lactic acidosis in metformin-treated patients, even when it is used in populations with relative contraindications. Some oral hypoglycemic medications can cause hypoglycemia. Hypoglycemia is especially common in older patients, alcoholics, and patients

with liver or renal disease. Patients on sulfonylureas and meglitinides have the highest incidence of hypoglycemia because of their pharmacological action of increasing insulin secretion. Of the sulfonylureas, glyburide presents the highest risk of hypoglycemia. Combination therapies, especially those regimens containing a sulfonylurea, increase the risk of hypoglycemia. *Mt Sinai J Med* 76:234–243, 2009. © 2009 Mount Sinai School of Medicine

Key Words: diabetes mellitus, hypoglycemia, lactic acidosis, oral hypoglycemic drugs.

Multiple treatment choices are available to control the hyperglycemia seen in patients with type 2 diabetes. Some of these have been the focus of media controversies which may affect patient compliance. We will therefore clarify the recent controversy surrounding the cardiovascular and bone-health risks of thiazolidinediones, the risk of lactic acidosis with metformin, and the risk of hypoglycemia with oral therapies.

SAFETY AND UTILITY OF THIAZOLIDINEDIONES

The thiazolidinediones are a group of oral hypoglycemic agents first introduced in 1997.¹ Their mechanism of action is not yet fully elucidated, but they activate peroxisome proliferator-activated receptor gamma (PPAR- γ) located within the nucleus of the cell, changing the transcription of more than 100 genes. Among these are genes that reduce hepatic insulin resistance and gluconeogenesis, thereby reducing glucose and insulin concentrations in the blood.^{2,3} The drugs have been popular because their mechanism of action complements other available oral therapies, and they have a low risk of hypoglycemia. These factors make them attractive choices for delaying insulin initiation. These agents

Address Correspondence to:

Graham T. McMahon

Division of Endocrinology
Diabetes, and Hypertension
Brigham and Women's Hospital
Boston, MA
Email: gmcmahon@partners.org

are also moderately effective, reducing hemoglobin A1c (HbA1c) levels by 0.7% to 1.6%^{4,5} in comparison with placebo treatment.

Heart Failure

The initial introduction of the thiazolidinediones was controversial because troglitazone, the first thiazolidinedione to be approved in 1997, was withdrawn in March 2000 after being linked with hepatotoxicity: 66 deaths and 11 liver transplants were reported to the US Food and Drug Administration (FDA) by the time it was withdrawn.⁶ Rosiglitazone and pioglitazone were both approved in 1999 and have not been associated with hepatotoxicity. It is now well recognized that rosiglitazone and pioglitazone are associated with fluid retention and an increased incidence of heart failure. In a retrospective cohort study performed in 2003, patients taking a thiazolidinedione had a 55% increased risk of heart failure in comparison with patients who did not take such agents.⁷ This prompted the FDA to add boxed warnings to the prescription information for the thiazolidinediones in August 2007.⁸

Cardiovascular Events

The most recently publicized controversy involving the thiazolidinediones was triggered by Nissen and Wolski's June 2007 meta-analysis implicating rosiglitazone in accelerating cardiovascular risk in patients with type 2 diabetes.⁹ The investigators aimed to assess the effect of rosiglitazone on cardiovascular morbidity and mortality through a meta-analysis of trials with outcome data for myocardial infarction and cardiac death. Trials eligible to be included in the meta-analysis were of more than 24 weeks' duration and included a comparison group not receiving rosiglitazone. Among the 42 included trials, the authors identified 86 myocardial infarctions in the rosiglitazone group and 72 in the control group [odds ratio (OR), 1.43; 95% confidence interval (CI), 1.03–1.98; $P = 0.03$] from among the 27,847 enrolled patients. The same trend was seen when total cardiovascular deaths were analyzed (39 deaths versus 22 deaths in the rosiglitazone and control groups: OR, 1.64; 95% CI, 0.98–2.74; $P = 0.06$). This additional toxicity was evident in short-term and longer term trials, suggesting an early adverse effect of rosiglitazone therapy.

Theories espoused by the authors to explain poor outcomes in thiazolidinedione-treated patients include raised low-density lipoprotein cholesterol, weight gain, congestive heart failure from increased intravascular volume, and reduced hemoglobin levels. The manufacturers and others characterized

the meta-analysis as flawed,¹⁰ citing the use of summary trial-level data rather than patient-level data (making it impossible to calculate more statistically powerful time-to-event analyses) and the inclusion of trials of widely varying sizes without a predefined method for adjudication of myocardial infarction or cardiovascular death. Finally, the total number of events was low, and this made the results vulnerable to change if only a few events had been misclassified.^{11,12} A subsequent (unpublished) study,¹³ sponsored by the manufacturer of rosiglitazone, similarly revealed that despite improvements in glycemic control and other metabolic variables with rosiglitazone, the expected reduction in cardiovascular events had not been found. The controversy prompted the release of an unplanned analysis of data from RECORD (Rosiglitazone Evaluated for Cardiac Outcomes and Regulation of Glycaemia in Diabetes).¹⁴ This study, the first to assess the effect of rosiglitazone on cardiovascular endpoints, did not show an increased risk of the primary endpoint of death or hospitalization from cardiovascular causes in patients receiving rosiglitazone: 217 patients in the rosiglitazone group had the adjudicated primary endpoint versus 202 in the control group (hazard ratio, 1.08; 95% CI, 0.89–1.31). These results have also been criticized: the primary outcome may be too nonspecific and therefore drive the relative risk toward null, and the trial was underpowered to show any difference in rates.¹⁵ Data from ADOPT (A Diabetes Outcome Progression Trial),¹⁶ a 4- to 6-year study of glycemic control in 4360 people recently diagnosed with type 2 diabetes and treated with rosiglitazone, metformin, or glibenclamide, was included in the meta-analysis by Nissen and Wolski.⁹ However, the data have since been reanalyzed by GlaxoSmithKline, and no increased risk with rosiglitazone was found.¹⁷ Data from a large observational study presented to the FDA by WellPoint (the largest health insurer in the United States) also did not reveal any increased risk in the rosiglitazone group.¹⁸ Furthermore, in ACCORD (Action To Control Cardiovascular Risk in Diabetes),¹⁹ no relationship was found between cardiovascular mortality and any particular drug, including rosiglitazone. Perhaps most troubling is the absence of any documented cardiovascular risk reduction with these drugs despite metabolic improvements.

A potential class effect of PPAR-receptor agonism was also implicated when the development of muraglitazar (a combined PPAR- α and PPAR- γ agonist) was halted after trials showed an increase in cardiovascular adverse events.²⁰ In contrast, pioglitazone has not been documented to produce a similarly increased risk of cardiovascular events. In

PROACTIVE (Prospective Pioglitazone Clinical Trial in Macrovascular Events), there was a trend toward a reduction in the primary endpoint (all-cause mortality, stroke, and coronary and peripheral vascular events) in the pioglitazone group (hazard ratio, 0.90; $P = 0.095$).²¹ Similar results for both pioglitazone and rosiglitazone were found by Gerrits *et al.*²²

Finally, data from publications using other endpoints such as C-reactive protein,²³ endothelial function,²⁴ the number and migratory activity of endothelial precursor cells,²⁵ restenosis after percutaneous coronary intervention,²⁶ and plaque stability²⁷ have all shown a beneficial effect of thiazolidinediones. An example is PERISCOPE (Pioglitazone Effect on Regression of Intravascular Sonographic Coronary Obstruction Prospective Evaluation).²⁸ In this study, 543 patients were randomized to receive glimepiride or pioglitazone for 18 months, and this was followed by coronary intravascular ultrasonography. The results were in favor of pioglitazone, with a 0.16% reduction of the percent atheroma volume in the pioglitazone group versus an increase of 0.73% in the glimepiride group.

This disparity in evidence led the FDA to decide against withdrawing rosiglitazone from the market at the recommendation of its advisory committee meeting in July 2007. It also led many commentators to appeal for appropriately nuanced media reporting, more extensive clinical trial evidence, and greater transparency from the manufacturers.²⁹

Bone Health

Patients with diabetes have an increased incidence of osteoporosis, but there is increasing evidence that thiazolidinediones accelerate bone loss even further and augment the risk of fracture.^{30–32} The PPAR receptors are widespread in human tissues, including bone. Early in vitro studies and animal models have shown that PPAR- γ receptors play a role in the differentiation of mesenchymal precursors in the bone marrow, favoring the formation of adipocytes over osteoblasts.^{33,34} Further studies have shown evidence of sustained reductions in markers of bone turnover³⁵ in thiazolidinedione-treated patients, whereas others have shown only a transient reduction. Clinical studies have examined the effect of thiazolidinediones on bone mineral density. Health ABC (Health, Aging, and Body Composition)³⁶ included 666 diabetic patients (70–79 years old), 69 of whom were taking a thiazolidinedione. Whole body bone loss was found to be 2.5 times higher per year in women taking a thiazolidinedione but not in men. However, men were found to have a similar decrease in bone mineral density

in another observational study.³⁷ A recent 14-week randomized controlled trial demonstrated that treatment with rosiglitazone was associated with a significantly greater loss of bone density at the hip (–1.9% and –0.2% in the rosiglitazone and placebo groups, respectively).³⁸ In the ADOPT trial,¹⁶ users of thiazolidinediones were shown to be at higher risk of foot and upper limb fractures than nonusers. The overall fracture rate in women was highest in patients treated with rosiglitazone (9.3% for rosiglitazone, 5.08% for metformin, and 3.47% for glyburide, $P < 0.01$); the rate of hip and spinal fractures did not vary between the groups. The distribution of risk factors for fracture did not differ between the 3 groups.³⁹ Similar outcomes were noted in pioglitazone-treated patients.⁴⁰ A recent observational study based on data from the United Kingdom General Practice Research Database provided further confirmation of the association.⁴¹ In patients who had taken thiazolidinediones for at least 12 to 18 months, the OR for fracture was 2.43 (95% CI, 1.49–3.95) after adjustments for other risk factors for fracture. No such association was found with other oral hypoglycemic medications.

Although thiazolidinediones effectively lower HbA1c and present a very low risk of hypoglycemia, they have not been convincingly associated with reductions in microvascular or macrovascular complications of diabetes. Evidence suggests that despite improvements in HbA1c, thiazolidinediones are associated with an increased risk of heart failure and may increase the risk of cardiovascular death. Furthermore, this class of drug appears to reduce bone density and has been associated with an increased risk of fracture. Given the availability of a wide array of alternative oral hypoglycemics,⁴² many clinicians are moving away from prescribing this class of drugs until further clinical trials such as RECORD and BARI 2D (Bypass Angioplasty Revascularization Investigation 2 Diabetes)⁴³ clarify their long-term clinical efficacy. Several agencies have decided to stop covering rosiglitazone, including the Department of Veterans Affairs, HealthTrans Pharmacy, and Prime Therapeutics.⁴⁴ Sales have dropped significantly: one study showed a fall in the number of claims for rosiglitazone from a managed care pharmacy from 98.8 per 1 million members in April 2007 to 31.8 per million members in May 2008.⁴⁴

METFORMIN

History

Biguanides (metformin, buformin, and phenformin) have been available since the late 1950s⁴⁵ and were,

in addition to the sulfonylureas, the first available oral hypoglycemic medications. They reduce fasting plasma glucose and insulin levels by reducing hepatic gluconeogenesis and improving glucose uptake in peripheral tissues.⁴⁶ Metformin reduces HbA1c by 1.3% to 2.0% and is associated with mild weight loss.⁴⁶ Transient mild gastrointestinal upset affects approximately 20% of patients,⁴⁷ but few patients need to discontinue therapy.

As early as in 1963,⁴⁸ metformin's predecessor phenformin was found to be associated with lactic acidosis. In the period prior to its withdrawal, 306 cases of phenformin-induced lactic acidosis were documented, with an overall mortality rate of 42%.⁴⁹ This led to the withdrawal of phenformin from the US market⁵⁰ in 1976. Because of similar concerns about metformin, it was not approved for use in the United States until 1995,⁵¹ although it was actively being used in Europe and other parts of the world during this time.

Although the risk of lactic acidosis with phenformin was generally not contested, the association between lactic acidosis and metformin has not been clearly established. This has led to an ongoing controversy as to whether metformin causes lactic acidosis and whether the strict prescribing guidelines are appropriate.

Mechanism

The underlying mechanism of lactic acidosis with biguanide use is unclear. Metformin has been proven to affect non-oxygen-dependent glucose utilization⁵² in the mitochondria,^{53,54} increasing lactate levels. This effect seems particularly pronounced at the level of the intestinal epithelium.⁵⁵ The lactate from the intestinal mucosa is generally cleared by the liver, although animal studies have shown a reduced uptake of lactate into the liver in the presence of metformin.⁵⁶ In human studies, the plasma level of lactate in metformin users seems to remain unchanged,^{57,58} although isolated cases have shown asymptomatic lactic acidemia⁵⁹ (this contrasts with phenformin, which causes a more marked increase in lactic acid levels⁶⁰).

Case Reports and Observational Studies of Metformin-Associated Lactic Acidosis

Patients with increasingly complex medical problems can develop lactic acidosis. Most of the reported cases of metformin-associated lactic acidosis have occurred in patients with other well-established causes of lactic acidosis.⁶¹ These reports include patients with acute renal failure (from dehydration,⁶²

radiological contrast agents,⁶³ nonsteroidal anti-inflammatory drugs or cyclooxygenase-2 inhibitors,⁶⁴ and angiotensin-converting enzyme inhibitors⁶⁵), alcoholics,⁶⁶ patients with liver failure,⁶⁷ and patients with heart failure⁶⁸ and other states of poor oxygen delivery to the tissues. As a result, current recommendations suggest that metformin be avoided in patients with a predisposition to tissue hypoxia, in patients with renal impairment (creatinine ≥ 1.5 for men or ≥ 1.4 for women), heart failure, respiratory failure, a history of lactic acidosis, sepsis causing reduced tissue perfusion, or liver disease, and in patients around the time of intravenous radiographic contrast agent use.⁴⁵ However, this excludes a large proportion of patients with diabetes—up to 50% in one study⁶⁹—some of whom have the most to gain from glucose control with metformin. Many clinicians are more liberal with the use of metformin than the guidelines recommend, and some have called for “metformin's contraindications [to be] contraindicated.”⁷⁰ Widespread awareness of these guidelines has given the perception that the association between metformin and lactic acidosis is well established and may have led to a publication bias.⁷¹

Metformin is cleared by the kidneys and accumulates in renal failure.^{46,72} Because renal failure may cause lactic acidosis, the elevated level of metformin seen in patients with lactic acidosis does not indicate a causative relationship. Elevated lactic acid levels have also occurred in patients who overdosed on metformin.^{73,74} There also has been no clear correlation between serum lactate levels and serum metformin levels.⁷⁵ Furthermore, the severity of the underlying comorbidity has been a significantly stronger predictor of outcome than the serum concentrations of metformin or lactate.^{71,74} However, the controversy persists as very occasional case reports⁷⁶ appear to describe lactic acidosis in otherwise healthy metformin recipients.

A key issue is whether metformin can cause lactic acidosis in patients without comorbidities. In the Freemantle Diabetes study, which included older patients with a high prevalence of comorbidities followed for a total of 17,638 patient years, the rate of lactic acidosis in metformin users was 57 per 100,000 patient years, which was not significantly different from the rate of 29 per 100,000 patient years noted in non-metformin users. In the year following metformin's approval in the United States, 47 cases of possible metformin-associated lactic acidosis were reported to the FDA.⁶¹ Forty-three of these had other risk factors for lactic acidosis, including heart failure, renal failure, chronic hypoxia, and old age. The reported rate of lactic acidosis was 5 cases

Table 1. Cases of Metformin-Associated Lactic Acidosis Reported from Population-Based Studies and Trials.

Country	Number of Cases	Number of Cases per 100,000 Patient Years	Number of Cases in Patients Without Comorbidities
United States ⁶¹	47	5	4
Canada ⁷⁷	2	9	0
Sweden ⁷⁸	16	15 for 1977–1981, 2.4 for 1987–1991	0
Australia ¹¹⁰	5	5.7	0
Systematic review of trial data ⁵⁸	0	5.1	0

per 100,000 patient years. This rate is similar to the rate of 2.4 to 15 cases per 100,000 patient years that has been reported in population-based studies from Saskatchewan⁷⁷ and Sweden⁷⁸ and similar to the rate of lactic acidosis in patients with type 2 diabetes in the United States prior to the introduction of metformin, which was estimated to be 9.7 to 16.9 cases per 100,000 patient years.⁷⁹ The incidence of lactic acidosis in clinical trials, which tend to recruit less complex patients, is substantially lower. A Cochrane systematic review⁵⁸ of 274 studies of metformin treatment found no cases of lactic acidosis after analyzing 59,321 patient years of metformin use. It estimated that the upper limit of the true incidence of lactic acidosis was 5.1 per 100,000 patient years. Because only 55% of the studies included in this meta-analysis excluded patients with renal impairment and 97% of the studies included patients with at least one of the major contraindications (renal insufficiency, cardiovascular disease, liver disease, or pulmonary disease), the true incidence of lactic acidosis in metformin-treated patients who do not have comorbidities may be significantly lower than could be extrapolated from the review. The rates of lactic acidosis in the main observational studies are indicated in Table 1. These results confirm that the incidence of lactic acidosis in patients without contributing comorbidities is likely to be very low.

To determine whether any additional cases clarified the relationship between metformin and lactic acidosis since the last meta-analysis, we performed a PubMed search and identified 44 cases published between September 2006 and August 2008 for which patient-level data were available (Table 2). Of the 44 cases, only 2 did not have an obvious trigger or contraindication. The mortality was 25% (11 cases).

Metformin Use in Populations with Contraindications

The use of metformin in patients with absolute or relative contraindications is common. In a study in Scotland of 1847 patients receiving metformin, 24.5% of patients were found to have a contraindication to metformin therapy.⁸⁰ Similar evidence of liberal use has been documented elsewhere.^{81,82} In a Scottish study, there was 1 episode of lactic acidosis corresponding to an incidence of 21.7 per 100,000 patient years; this was significantly higher than the rates noted previously but similar to the rates noted in the Fremantle Diabetes Study. This finding was not replicated in the other studies.

Metformin and Acidosis

The scarcity of evidence for definitively establishing an association between metformin and lactic

Table 2. Overview of Case Reports Published from September 2006 to August 2008 (Inclusive).

Contraindication/Trigger	Number of Cases	References
Acute renal failure from dehydration	20	62, 65, 68, 111, 112
Used in a patient with chronic renal disease/on dialysis	6	73, 110, 113
Medication overdose	5	73, 114–116
Acute exacerbation of congestive heart failure or acute myocardial infarction	3	68, 110
Acute renal failure from medications	3	64, 76, 117
Acute renal failure, multifactorial or of unknown etiology	2	111, 118
Acute renal failure from a radiological contrast agent	1	63
Shock	1	110
Multiple risk factors for lactic acidosis	1	119
No obvious contraindication or triggering factor	2	76, 120
Deaths	11	68, 73, 110, 111, 113, 118

Table 3. *Comparative Incidence of Selected Severe Adverse Events Associated with Oral Hypoglycemic Drugs.*

Complication	Estimated Incidence per 100,000 Patient Years
Lactic acidosis in metformin users	Up to 57
Heart failure in thiazolidinedione users ⁷	870
Severe hypoglycemia in sulfonylurea users ¹²¹	2402

acidosis contrasts sharply with the proven benefits of metformin therapy, even in patients with comorbidities.^{83,84} The recent publication of follow-up data from the United Kingdom Prospective Diabetes Study showed that participants in the metformin-treated group had a 33% reduced risk of myocardial infarction and a 27% reduced risk of death from any cause 10 years after the termination of the study.⁸⁵ In view of its proven safety and efficacy and its potential for cardiovascular benefit, the joint American Diabetes Association and European Association for the Study of Diabetes consensus guidelines on the management of hyperglycemia in type 2 diabetes⁸⁶ explicitly state that metformin should be used as first-line therapy, in addition to lifestyle modification. The low incidence of lactic acidosis compares favorably with risks associated with other oral hypoglycemic medications (Table 3).

It has thus been suggested that restrictions should be less broad,⁸⁷ allowing metformin to be used in patients with stable heart failure,⁸⁸ stable renal impairment,⁸⁹ and advanced age. Although data support its safety in such patients, evidence from the case studies detailed here suggest that it is optimal to discontinue metformin in the setting of acute disease states in which there is a risk of tissue hypoxia or renal failure, such as sepsis or acute myocardial infarction, and perioperatively.⁹⁰

HYPOGLYCEMIA

Systematic reviews have consistently demonstrated the increased risk of hypoglycemia associated with the use of sulfonylureas. This risk is directly attributable to the pharmacological action of this class of drugs, which augment insulin secretion even when glucose levels are low. Reported incidence rates for hypoglycemia vary widely; however, the absolute risk of hypoglycemia is 4% to 9% higher in sulfonylurea-treated patients versus patients treated with thiazolidinediones or metformin.⁹¹ The rate of hypoglycemia in patients treated with a meglitinide

is largely similar to that with other sulfonylureas.⁹¹ Inhibitors of dipeptidyl peptidase 4 such as sitagliptin, vildagliptin, and saxagliptin are associated with a very low risk of hypoglycemia.^{92,93}

Glyburide is associated with the highest risk of hypoglycemia, with a lower risk of hypoglycemia in patients treated with glimepiride and glipizide.^{94–98} Because the metabolites of glyburide are metabolically active and accumulate in patients with renal insufficiency, hypoglycemia rates are especially high and hypoglycemia events are prolonged when glyburide is used to treat diabetes in patients with impaired renal function.⁹⁹ In contrast, glipizide and glimepiride less frequently cause hypoglycemia as their metabolites are largely inactive.⁹⁹

Hypoglycemia is especially problematic in the older patient, particularly in the month after hospital discharge.¹⁰⁰ Other clinical factors associated with drug-induced hypoglycemia are alcohol use and liver or renal disease.¹⁰¹ Risk factors in combination significantly augment that risk.¹⁰¹

Combination therapy capitalizes on the opportunity to address defects in insulin secretion and insulin resistance in patients with diabetes. Hypoglycemia is the dominant adverse event seen with increasing combinations of oral hypoglycemic drugs, predominantly when sulfonylureas are part of the combination. Polypharmacy with additional drugs increases the risk for drug interactions.¹⁰⁰ Drugs that inhibit the cytochrome P450 metabolic pathway increase the risk of hypoglycemia. Known potentiators include salicylates, pyrazolone derivatives, monoamine oxidase inhibitors, sulfonamides, coumarin anticoagulants, chloramphenicol, and clofibrate.¹⁰² Antihypertensives do not appear to increase the risk of hypoglycemia.¹⁰³

Combining certain oral hypoglycemic drugs with insulin can reduce the amount of weight gain and reduce the total daily insulin dose required for glycemic efficacy.^{104–107} In general, treatments that combine oral hypoglycemic drugs with insulin lead to a lower incidence of hypoglycemia than insulin treatment alone, likely as a result of the reduced insulin dose needed when oral medications are coprescribed.^{104–107} Sulfonylureas are associated with a significantly higher rate of hypoglycemia than metformin in insulin-treated patients and should therefore generally be withheld as insulin is initiated.¹⁰⁸

There are other oral medications available with a lower risk of hypoglycemia than the sulfonylureas. Acarbose is an alpha-glucosidase inhibitor that acts locally in the gastrointestinal tract to delay the absorption of carbohydrates, thus causing less postprandial hyperglycemia. It can reduce HbA1c by 0.5% to 1%, but its use is often limited by gastrointestinal

Table 4. *Key Points.*

Rosiglitazone may be associated with an increased risk of cardiovascular events.
Thiazolidinediones appear to increase the risk of fractures and heart failure.
The risk of lactic acidosis with metformin is very low if it is used appropriately.
Sulfonylureas are the oral antidiabetic agents most associated with a risk of hypoglycemia.

side effects such as bloating, abdominal discomfort, flatulence, and diarrhea.¹⁰⁹ Newer drugs, such as inhibitors of dipeptidyl peptidase 4, appear to be associated with a low incidence of adverse events, including hypoglycemia. However, the ubiquity of the enzyme in diverse organ systems and the short duration of clinical trials and real-world experience are reasons to be cautious.

CONCLUSION

There is a wide array of oral medications available to manage the hyperglycemia associated with type 2 diabetes mellitus (Table 4). Metformin is the first choice as suggested by the American Diabetes Association. It is safe in most patients and presents a low risk of weight gain or hypoglycemia. Sulfonylureas remain the standard choice for add-on therapy, although the risk of hypoglycemia needs to be considered, especially in the elderly and in patients with renal or liver failure. Acarbose is safe, but its use is often limited by its side effects.

The newer agents, including the thiazolidinediones and the dipeptidyl peptidase 4 inhibitors, still have an unclear safety profile and are still considered second or third-line therapy. Insulin initiation may be a safer and more validated therapy when double-agent therapy fails. Insulin initiation is also more appropriate late in the course of type 2 diabetes when insulin secretagogues (sulfonylureas and meglitinides) are ineffective because of beta-cell exhaustion, which makes further stimulation of endogenous insulin secretion impossible.

It has become clear that when a wide variety of effective drugs are available, clinicians must increasingly be cognizant of the adverse event profile of the drugs that they prescribe and recognize the unique patient characteristics that make rational prescribing most appropriate.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. US Food and Drug Administration. Approvals for January 1997. Available at: <http://www.fda.gov/cder/da/da0197.htm>. Accessed April 2009.
2. Campbell IW. The clinical significance of PPAR gamma agonism. *Curr Mol Med* 2005; 5: 349–363.
3. Yki-Järvinen H. Thiazolidinediones. *N Engl J Med* 2004; 351: 1106–1118.
4. Aronoff S, Rosenblatt S, Braithwaite S, et al. Pioglitazone hydrochloride monotherapy improves glycemic control in the treatment of patients with type 2 diabetes: a 6-month randomized placebo-controlled dose-response study. The Pioglitazone 001 Study Group. *Diabetes Care* 2000; 23: 1605–1611.
5. Lebovitz HE, Dole JF, Patwardhan R, et al. Rosiglitazone monotherapy is effective in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2001; 86: 280–288.
6. Gale EA. Troglitazone: the lesson that nobody learned? *Diabetologia* 2006; 49: 1–6.
7. Delea TE, Edelsberg JS, Hagiwara M, et al. Use of thiazolidinediones and risk of heart failure in people with type 2 diabetes: a retrospective cohort study. *Diabetes Care* 2003; 26: 2983–2989.
8. US Food and Drug Administration. MedWatch August 2007. Safety-related drug labeling changes—detailed. Available at: <http://www.fda.gov/medwatch/safety/2007/aug07.htm>. Accessed April 2009.
9. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; 356: 2457–2471.
10. McGuire DK. Rosiglitazone and cardiovascular disease: an epidemiologist's perspective. *Diab Vasc Dis Res* 2007; 4: 77–79.
11. Psaty BM, Furberg CD. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; 356: 2522–2524.
12. Bracken MB. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; 357: 937–938.
13. GlaxoSmithKline. Study no. ZM2005/00181/01: Avandia cardiovascular event modeling project. Available at: <http://www.gsk-clinicalstudyregister.com/files/pdf/24040.pdf>. Accessed April 2009.
14. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes—an interim analysis. *N Engl J Med* 2007; 357: 28–38.
15. Psaty BM, Furberg CD. The record on rosiglitazone and the risk of myocardial infarction. *N Engl J Med* 2007; 357: 67–69.
16. Kahn SE, Haffner SM, Heise MA, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *N Engl J Med* 2006; 355: 2427–2443.
17. Krall RL. Cardiovascular safety of rosiglitazone. *Lancet* 2007; 369: 1995–1996.
18. Rosen CJ. The rosiglitazone story—lessons from an FDA advisory committee meeting. *N Engl J Med* 2007; 357: 844–846.
19. Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 2008; 358: 2545–2559.
20. Nissen SE, Wolski K, Topol EJ. Effect of muraglitazar on death and major adverse cardiovascular events in patients with type 2 diabetes mellitus. *JAMA* 2005; 294: 2581–2586.

21. Dormandy JA, Charbonnel B, Eckland DJ, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROACTIVE Study (Prospective Pioglitazone Clinical Trial in Macrovascular Events): a randomised controlled trial. *Lancet* 2005; 366: 1279–1289.
22. Gerrits CM, Bhattacharya M, Manthena S, et al. A comparison of pioglitazone and rosiglitazone for hospitalization for acute myocardial infarction in type 2 diabetes. *Pharmacoepidemiol Drug Saf* 2007; 16: 1065–1071.
23. Haffner SM, Greenberg AS, Weston WM, et al. Effect of rosiglitazone treatment on nontraditional markers of cardiovascular disease in patients with type 2 diabetes mellitus. *Circulation* 2002; 106: 679–684.
24. Natali A, Toschi E, Baldeweg S, et al. Haematocrit, type 2 diabetes, and endothelium-dependent vasodilatation of resistance vessels. *Eur Heart J* 2005; 26: 464–471.
25. Pistrosch F, Herbrig K, Oelschlaegel U, et al. PPARgamma-agonist rosiglitazone increases number and migratory activity of cultured endothelial progenitor cells. *Atherosclerosis* 2005; 183: 163–167.
26. Choi D, Kim SK, Choi SH, et al. Preventative effects of rosiglitazone on restenosis after coronary stent implantation in patients with type 2 diabetes. *Diabetes Care* 2004; 27: 2654–2660.
27. Marfella R, D'Amico M, Esposito K, et al. The ubiquitin-proteasome system and inflammatory activity in diabetic atherosclerotic plaques: effects of rosiglitazone treatment. *Diabetes* 2006; 55: 622–632.
28. Nissen SE, Nicholls SJ, Wolski K, et al. Comparison of pioglitazone vs glimepiride on progression of coronary atherosclerosis in patients with type 2 diabetes: the PERISCOPE randomized controlled trial. *JAMA* 2008; 299: 1561–1573.
29. Marx N. Rosiglitazone and cardiovascular disease: a cardiologist's perspective. *Diab Vasc Dis Res* 2007; 4: 80–81.
30. Watanabe S, Takeuchi Y, Fukumoto S, et al. Decrease in serum leptin by troglitazone is associated with preventing bone loss in type 2 diabetic patients. *J Bone Miner Metab* 2003; 21: 166–171.
31. Okazaki R, Totsuka Y, Hamano K, et al. Metabolic improvement of poorly controlled noninsulin-dependent diabetes mellitus decreases bone turnover. *J Clin Endocrinol Metab* 1997; 82: 2915–2920.
32. Meyer HE, Tverdal A, Falch JA. Risk factors for hip fracture in middle-aged Norwegian women and men. *Am J Epidemiol* 1993; 137: 1203–1211.
33. Akune T, Ohba S, Kamekura S, et al. PPARgamma insufficiency enhances osteogenesis through osteoblast formation from bone marrow progenitors. *J Clin Invest* 2004; 113: 846–855.
34. Rzonca SO, Suva LJ, Gaddy D, et al. Bone is a target for the antidiabetic compound rosiglitazone. *Endocrinology* 2004; 145: 401–406.
35. Okazaki R, Miura M, Toriumi M, et al. Short-term treatment with troglitazone decreases bone turnover in patients with type 2 diabetes mellitus. *Endocr J* 1999; 46: 795–801.
36. Schwartz AV, Sellmeyer DE, Vittinghoff E, et al. Thiazolidinedione use and bone loss in older diabetic adults. *J Clin Endocrinol Metab* 2006; 91: 3349–3354.
37. Yaturu S, Bryant B, Jain SK. Thiazolidinedione treatment decreases bone mineral density in type 2 diabetic men. *Diabetes Care* 2007; 30: 1574–1576.
38. Grey A, Bolland M, Gamble G, et al. The peroxisome proliferator-activated receptor-gamma agonist rosiglitazone decreases bone formation and bone mineral density in healthy postmenopausal women: a randomized, controlled trial. *J Clin Endocrinol Metab* 2007; 92: 1305–1310.
39. Kahn SE, Zinman B, Lachin JM, et al. Rosiglitazone-associated fractures in type 2 diabetes: an analysis from A Diabetes Outcome Progression Trial (ADOPT). *Diabetes Care* 2008; 31: 845–851.
40. Takeda Pharmaceuticals North America, Inc. Re: Observation of an increased incidence of fractures in female patients who received long-term treatment with ACTOS®(pioglitazone HCl) tablets for type 2 diabetes mellitus. Available at: <http://www.fda.gov/medwatch/safety/2007/actosmar0807.pdf>. Accessed April 2009.
41. Meier C, Kraenzlin ME, Bodmer M, et al. Use of thiazolidinediones and fracture risk. *Arch Intern Med* 2008; 168: 820–825.
42. Nathan DM. Rosiglitazone and cardiotoxicity—weighing the evidence. *N Engl J Med* 2007; 357: 64–66.
43. Magee MF, Isley WL. Rationale, design, and methods for glycemic control in the Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) Trial. *Am J Cardiol* 2006; 97: 20G–30G.
44. Starner CI, Schafer JA, Heaton AH, Gleason PP. Rosiglitazone and pioglitazone utilization from January 2007 through May 2008 associated with five risk-warning events. *J Manag Care Pharm* 2008; 14: 523–531.
45. Bailey CJ, Turner RC. Metformin. *N Engl J Med* 1996; 334: 574–579.
46. Kirpichnikov D, McFarlane SI, Sowers JR. Metformin: an update. *Ann Intern Med* 2002; 137: 25–33.
47. DeFronzo RA, Goodman AM. Efficacy of metformin in patients with non-insulin-dependent diabetes mellitus. The Multicenter Metformin Study Group. *N Engl J Med* 1995; 333: 541–549.
48. Tranquada RE, Bernstein S, Martin HE. Irreversible lactic acidosis associated with phenformine therapy. Report of three cases. *JAMA* 1963; 184: 37–42.
49. Misbin RI. Phenformin-associated lactic acidosis: pathogenesis and treatment. *Ann Intern Med* 1977; 87: 591–595.
50. US Food and Drug Administration. Why drugs get pulled off the market. Available at: http://www.fda.gov/fdac/features/2002/102_drug.html. Accessed April 2009.
51. US Food and Drug Administration. Orange book detail record search. Available at: http://www.accessdata.fda.gov/scripts/cder/ob/docs/obdetail.cfm?appl_no=020357&table1=ob_rx. Accessed April 2009.
52. Bailey CJ. Biguanides and NIDDM. *Diabetes Care* 1992; 15: 755–772.
53. Steiner DF, Williams RH. Respiratory inhibition and hypoglycemia by biguanides and decamethylenediguanidine. *Biochim Biophys Acta* 1958; 30: 329–340.
54. Owen MR, Doran E, Halestrap AP. Evidence that metformin exerts its anti-diabetic effects through

- inhibition of complex I of the mitochondrial respiratory chain. *Biochem J* 2000; 348(pt 3): 607–614.
55. Bailey CJ, Wilcock C, Day C. Effect of metformin on glucose metabolism in the splanchnic bed. *Br J Pharmacol* 1992; 105: 1009–1013.
56. Radziuk J, Zhang Z, Wiernsperger N, Pye S. Effects of metformin on lactate uptake and gluconeogenesis in the perfused rat liver. *Diabetes* 1997; 46: 1406–1413.
57. Rachmani R, Slavachevski I, Levi Z, et al. Metformin in patients with type 2 diabetes mellitus: reconsideration of traditional contraindications. *Eur J Intern Med* 2002; 13: 428.
58. Salpeter S, Greyber E, Pasternak G, Salpeter E. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2006; CD002967.
59. Khan JK, Pallaki M, Tolbert SR, Hornick TR. Lactic acidemia associated with metformin. *Ann Pharmacother* 2003; 37: 66–69.
60. Cavallo-Perin P, Aluffi E, Estivi P, et al. The hyperlactatemic effect of biguanides: a comparison between phenformin and metformin during a 6-month treatment. *Riv Eur Sci Med Farmacol* 1989; 11: 45–49.
61. Misbin RI, Green L, Stadel BV, et al. Lactic acidosis in patients with diabetes treated with metformin. *N Engl J Med* 1998; 338: 265–266.
62. El-Hennawy AS, Jacob S, Mahmood AK. Metformin-associated lactic acidosis precipitated by diarrhea. *Am J Ther* 2007; 14: 403–405.
63. Jain V, Sharma D, Prabhakar H, Dash HH. Metformin-associated lactic acidosis following contrast media-induced nephrotoxicity. *Eur J Anaesthesiol* 2008; 25: 166–167.
64. Tewari P, Nath SS, Mazumdar G. Increased prothrombin time and lactic acidosis: rofecoxib drug interaction with acenocoumarin and metformin. *Ann Card Anaesth* 2007; 10: 58–60.
65. Gudmundsdottir H, Aksnes H, Heldal K, et al. Metformin and antihypertensive therapy with drugs blocking the renin angiotensin system, a cause of concern? *Clin Nephrol* 2006; 66: 380–385.
66. Houwerzijl EJ, Snoek WJ, van Haastert M, Holman ND. Severe lactic acidosis due to metformin therapy in a patient with contra-indications for metformin [in Dutch]. *Ned Tijdschr Geneesk* 2000; 144: 1923–1926.
67. Edwards CM, Barton MA, Snook J, et al. Metformin-associated lactic acidosis in a patient with liver disease. *QJM* 2003; 96: 315–316.
68. Almirall J, Briculle M, Gonzalez-Clemente JM. Metformin-associated lactic acidosis in type 2 diabetes mellitus: incidence and presentation in common clinical practice. *Nephrol Dial Transplant* 2008; 23: 2436–2438.
69. Sulkin TV, Bosman D, Krentz AJ. Contraindications to metformin therapy in patients with NIDDM. *Diabetes Care* 1997; 20: 925–928.
70. McCormack J, Johns K, Tildesley H. Metformin's contraindications should be contraindicated. *CMAJ* 2005; 173: 502–504.
71. Stades AM, Heikens JT, Erkelens DW, et al. Metformin and lactic acidosis: cause or coincidence? A review of case reports. *J Intern Med* 2004; 255: 179–187.
72. Lalau JD, Lacroix C, Compagnon P, et al. Role of metformin accumulation in metformin-associated lactic acidosis. *Diabetes Care* 1995; 18: 779–784.
73. Perrone J, Phillips C, Gaieski D. Occult metformin toxicity in three patients with profound lactic acidosis. *J Emerg Med* 2008.
74. Lalau JD, Mourlhon C, Bergeret A, Lacroix C. Consequences of metformin intoxication. *Diabetes Care* 1998; 21: 2036–2037.
75. Lalau JD, Lacroix C, De Cagny B, Fournier A. Metformin-associated lactic acidosis in diabetic patients with acute renal failure. A critical analysis of its pathogenesis and prognosis. *Nephrol Dial Transplant* 1994; 9(suppl 4): 126–129.
76. Silvestre J, Carvalho S, Mendes V, et al. Metformin-induced lactic acidosis: a case series. *J Med Case Rep* 2007; 1: 126.
77. Stang M, Wysowski DK, Butler-Jones D. Incidence of lactic acidosis in metformin users. *Diabetes Care* 1999; 22: 925–927.
78. Wiholm BE, Myrhed M. Metformin-associated lactic acidosis in Sweden 1977–1991. *Eur J Clin Pharmacol* 1993; 44: 589–591.
79. Brown JB, Pedula K, Barzilay J, et al. Lactic acidosis rates in type 2 diabetes. *Diabetes Care* 1998; 21: 1659–1663.
80. Emslie-Smith AM, Boyle DI, Evans JM, et al. Contraindications to metformin therapy in patients with type 2 diabetes—a population-based study of adherence to prescribing guidelines. *Diabet Med* 2001; 18: 483–488.
81. Calabrese AT, Coley KC, DaPos SV, et al. Evaluation of prescribing practices: risk of lactic acidosis with metformin therapy. *Arch Intern Med* 2002; 162: 434–437.
82. Horlen C, Malone R, Bryant B, et al. Frequency of inappropriate metformin prescriptions. *JAMA* 2002; 287: 2504–2505.
83. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* 1998; 352: 854–865.
84. Masoudi FA, Inzucchi SE, Wang Y, et al. Thiazolidinediones, metformin, and outcomes in older patients with diabetes and heart failure: an observational study. *Circulation* 2005; 111: 583–590.
85. Holman RR, Paul SK, Bethel MA, et al. Long-term follow-up after tight control of blood pressure in type 2 diabetes. *N Engl J Med* 2008; 359: 1565–1576.
86. Nathan DM, Buse JB, Davidson MB, et al. Management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* 2006; 29: 1963–1972.
87. Holstein A, Stumvoll M. Contraindications can damage your health—is metformin a case in point? *Diabetologia* 2005; 48: 2454–2459.
88. MacDonald MR, Petrie MC, Hawkins NM, et al. Diabetes, left ventricular systolic dysfunction, and chronic heart failure. *Eur Heart J* 2008; 29: 1224–1240.
89. Herrington WG, Levy JB. Metformin: effective and safe in renal disease? *Int Urol Nephrol* 2008; 40: 411–417.
90. Vreven R, De Kock M. Metformin lactic acidosis and anaesthesia: myth or reality? *Acta Anaesthesiol Belg* 2005; 56: 297–302.

91. Bolen S, Feldman L, Vassy J, et al. Systematic review: comparative effectiveness and safety of oral medications for type 2 diabetes mellitus. *Ann Intern Med* 2007; 147: 386–399.
92. Gagliardino JJ, Santoro S, Arellano S, Di Girolamo G. New treatments for type 2 diabetes mellitus: combined therapy with sitagliptin. *Expert Opin Pharmacother* 2008; 9: 1495–1507.
93. Scott R, Loeys T, Davies MJ, Engel SS. Efficacy and safety of sitagliptin when added to ongoing metformin therapy in patients with type 2 diabetes. *Diabetes Obes Metab* 2008; 10: 959–969.
94. Campbell RK. Glimepiride: role of a new sulfonylurea in the treatment of type 2 diabetes mellitus. *Ann Pharmacother* 1998; 32: 1044–1052.
95. Dills DG, Schneider J. Clinical evaluation of glimepiride versus glyburide in NIDDM in a double-blind comparative study. Glimepiride/Glyburide Research Group. *Horm Metab Res* 1996; 28: 426–429.
96. Holstein A, Plaschke A, Egberts EH. Lower incidence of severe hypoglycaemia in patients with type 2 diabetes treated with glimepiride versus glibenclamide. *Diabetes Metab Res Rev* 2001; 17: 467–473.
97. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Individual sulfonylureas and serious hypoglycemia in older people. *J Am Geriatr Soc* 1996; 44: 751–755.
98. Jonsson A, Rydberg T, Ekberg G, et al. Slow elimination of glyburide in NIDDM subjects. *Diabetes Care* 1994; 17: 142–145.
99. Harrower AD. Pharmacokinetics of oral antihyperglycaemic agents in patients with renal insufficiency. *Clin Pharmacokinet* 1996; 31: 111–119.
100. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Incidence and risk factors for serious hypoglycemia in older persons using insulin or sulfonylureas. *Arch Intern Med* 1997; 157: 1681–1686.
101. Seltzer HS. Drug-induced hypoglycemia. A review of 1418 cases. *Endocrinol Metab Clin North Am* 1989; 18: 163–183.
102. van Staa T, Abenhaim L, Monette J. Rates of hypoglycemia in users of sulfonylureas. *J Clin Epidemiol* 1997; 50: 735–741.
103. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Antihypertensives and the risk of serious hypoglycemia in older persons using insulin or sulfonylureas. *JAMA* 1997; 278: 40–43.
104. Chow CC, Tsang LW, Sorensen JP, Cockram CS. Comparison of insulin with or without continuation of oral hypoglycemic agents in the treatment of secondary failure in NIDDM patients. *Diabetes Care* 1995; 18: 307–314.
105. Goudswaard AN, Stolk RP, Zuithoff P, et al. Starting insulin in type 2 diabetes: continue oral hypoglycemic agents? A randomized trial in primary care. *J Fam Pract* 2004; 53: 393–399.
106. Kabadi MU, Kabadi UM. Efficacy of sulfonylureas with insulin in type 2 diabetes mellitus. *Ann Pharmacother* 2003; 37: 1572–1576.
107. Kabadi UM, Kabadi M. Comparative efficacy of glimepiride and/or metformin with insulin in type 2 diabetes. *Diabetes Res Clin Pract* 2006; 72: 265–270.
108. Yki-Jarvinen H, Ryysy L, Nikkila K, et al. Comparison of bedtime insulin regimens in patients with type 2 diabetes mellitus. A randomized, controlled trial. *Ann Intern Med* 1999; 130: 389–396.
109. Cheng AY, Fantus IG. Oral antihyperglycemic therapy for type 2 diabetes mellitus. *CMAJ* 2005; 172: 213–226.
110. Kamber N, Davis WA, Bruce DG, Davis TM. Metformin and lactic acidosis in an Australian community setting: the Fremantle Diabetes Study. *Med J Aust* 2008; 188: 446–449.
111. Bruijstens LA, van Luin M, Buscher-Jungerhans PM, Bosch FH. Reality of severe metformin-induced lactic acidosis in the absence of chronic renal impairment. *Neth J Med* 2008; 66: 185–190.
112. Kumar U, Chennavir B, Gopi A. Heeding clues to metformin-associated lactic acidosis: prompt response can save life. *Emerg Med J* 2008; 25: 537–538.
113. Runge S, Mayerle J, Warnke C, et al. Metformin-associated lactic acidosis in patients with renal impairment solely due to drug accumulation? *Diabetes Obes Metab* 2008; 10: 91–93.
114. Dell'aglio DM, Perino LJ, Todino JD, et al. Metformin overdose with a resultant serum pH of 6.59: survival without sequelae. *J Emerg Med* 2008.
115. Galea M, Jelacin N, Bramham K, White I. Severe lactic acidosis and rhabdomyolysis following metformin and ramipril overdose. *Br J Anaesth* 2007; 98: 213–215.
116. Hong YC, O'Boyle CP, Chen IC, et al. Metformin-associated lactic acidosis in a pregnant patient. *Gynecol Obstet Invest* 2008; 66: 138–141.
117. Audia P, Feinfeld DA, Dubrow A, Winchester JF. Metformin-induced lactic acidosis and acute pancreatitis precipitated by diuretic, celecoxib, and candesartan-associated acute kidney dysfunction. *Clin Toxicol* 2008; 46: 164–166.
118. Gambaro V, Dell'acqua L, Fare F, et al. A case of fatal intoxication from metformin. *J Forensic Sci* 2007; 52: 988–991.
119. Riesenman PJ, Braithwaite SS, Cairns BA. Metformin-associated lactic acidosis in a burn patient. *J Burn Care Res* 2007; 28: 342–347.
120. Packer CD. Metformin-associated lactic acidosis in a patient with vertebral artery dissection. *South Med J* 2006; 99: 1147–1148.
121. Asche CV, McAdam-Marx C, Shane-McWhorter L, et al. Evaluation of adverse events of oral antihyperglycemic monotherapy experienced by a geriatric population in a real-world setting: a retrospective cohort analysis. *Drugs Aging* 2008; 25: 611–622.

Targeting Incretins in Type 2 Diabetes Mellitus

Matthew Potenza, MD, and Elliot J. Rayfield, MD

Mount Sinai School of Medicine, New York, NY

ABSTRACT

The incretins have emerged as key targets in the modern treatment of type 2 diabetes mellitus. Understanding the physiology of the incretins is essential to the physician's ability to appropriately use emerging pharmacotherapies that target this system. This review describes incretin physiology and discusses recent trials of drugs that modulate this system in the treatment of type 2 diabetes. A MEDLINE search using the terms "GLP-1" (ie, glucagon-like peptide 1), "incretins," "exenatide," and "DPP-IV inhibitors" (ie, dipeptidyl peptidase IV inhibitors) was performed, and pertinent articles from the past 10 years were reviewed. Articles describing incretin physiology and clinical trials with exenatide and dipeptidyl peptidase IV inhibitors were identified and discussed. As the articles show, new medications manipulating the incretin system are an important part of treating type 2 diabetes. The cost of these drugs and their potential side effects in comparison with existing agents must be considered when they are being selected as part of a treatment regimen. However, the evidence to date offers much promise and enthusiasm. *Mt Sinai J Med* 76:244–256, 2009. © 2009 Mount Sinai School of Medicine

Key Words: incretins, pharmacotherapy, type 2 diabetes mellitus.

Type 2 diabetes mellitus is a growing epidemic worldwide, and this is directly correlated to the increased prevalence of obesity, which has doubled in the past 30 years.¹ Obesity increases the risk of diabetes 90-fold; furthermore, obese individuals have a demonstrated abnormal glucose metabolism in comparison with their thinner counterparts, including

insulin resistance with resulting hyperinsulinemia, hyperglycemia, and dyslipidemia.² Loss of abdominal obesity has been shown to prevent the onset of diabetes and improve outcomes in patients with type 2 diabetes.³

Abdominal visceral fat contributes to insulin resistance and the development of type 2 diabetes. Mechanisms include increased production of proinflammatory cytokines by adipose tissue as well as increased release of free fatty acids, both of which worsen insulin resistance.⁴ Adipokines, or cytokines produced by visceral fat, have demonstrated associations with the development of type 2 diabetes and coronary artery disease.^{5,6} Weight is typically lost from visceral fat rather than subcutaneous fat. Thus, the benefit of weight loss in preventing and treating diabetes is likely due to altered production of these inflammatory molecules as visceral fat mass decreases.

Most medical treatments for diabetes, including insulin, insulin secretagogues, and thiazolidinediones (TZDs), paradoxically result in weight gain. TZDs do have a benefit, in that they have been shown to redistribute fat in some studies, moving the harmful visceral fat to less harmful subcutaneous fat.^{7–9} Acarbose, an inhibitor of intestinal alpha-glucosidase that prevents glucose absorption, and metformin, an insulin sensitizer, were the only medications to demonstrate weight loss and diabetes prevention before the introduction of the incretins.^{10–12} The incretins are peptides synthesized in the intestinal tract that have provided a new approach to treating and potentially preventing type 2 diabetes. The 2 main incretins are glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide 1 (GLP-1). This review focuses primarily on the manipulation of GLP-1 via the administration of its receptor agonists or via the inhibition of its degradation. GIP has not been investigated as a therapeutic agent as extensively as GLP-1, as discussed further later.

Address Correspondence to:

Matthew Potenza, MD

Mount Sinai School of Medicine
New York, NY

Email: matthew.potenza@mssm.edu

PHYSIOLOGY OF INCRETINS

The incretin effect refers to increased insulin release from the beta-cells of the pancreas in response to oral glucose and fat ingestion that is lost in intravenous glucose administration.¹³ Somehow, food entering the gut signals the pancreas to release more insulin and less glucagon. The incretin effect linking the gut and the pancreas was initially observed in experiments measuring curves of insulin response to oral glucose administration versus intravenous glucose administration.^{14,15} Greater amounts of insulin are released when nutrients are administered orally rather than intravenously, and glucose homeostasis is achieved more rapidly.

The incretin effect is distinct from but likely related to the first-phase insulin response. Intravenous glucose challenges elicit a first-phase insulin release from the pancreas within 0 to 8 minutes, which then is followed by a sustained delayed insulin response that is responsible for ultimate glucose homeostasis. The initial insulin response is blunted 3- to 4-fold in diabetic subjects, and this suggests that restoration of this first response may be important when type 2 diabetes is being treated.¹⁶ The first-phase insulin response also augments the second phase, so when first-phase insulin release is lost, the second-phase release is also blunted. The measurement of first-phase insulin response has become a marker of beta-cell function.¹⁷

Investigators have observed that infusing GIP or GLP-1 into patients with type 2 diabetes recovers the first-phase insulin response, and this suggests an improvement in beta-cell function. Postprandial levels of GLP-1 are markedly reduced in obese

individuals and in patients with type 2 diabetes in comparison with their nonobese, nondiabetic peers.¹⁸ The deficit of the incretin hormones, the loss of first-phase insulin release, and the recovery of the first phase with GLP-1 infusion suggest a link between incretins and first-phase insulin release. Although this is not demonstrated in oral glucose administration, it is possible that the enhanced insulin release seen with the incretin effect is related to the beta-cell first-phase release of preformed insulin-containing granules. Because the recovery of first-phase insulin response augments the second phase of sustained insulin release, it is believed that glucose homeostasis may be greatly enhanced by incretin therapy.^{19,20} The deficit of this hormone together with its apparent effects on insulin secretion makes GLP-1 an attractive target of potential therapy for type 2 diabetes. Replacement of the GLP-1 hormone could restore beta-cell function and even reverse the pathophysiology that causes type 2 diabetes.

One of the most notable advantages of GLP-1 as a therapeutic agent is the characteristic glucose-dependent nature of the incretin effect; this means that the incretins enhance insulin secretion only in the presence of glucose and fat in the gut (see Figure 1). This nutrient-dependent insulin release greatly reduces hypoglycemia.²¹ The molecular mechanisms of this nutrient-dependent effect have been explored by a number of investigators putting forth various hypotheses.^{22–25} Neuronal, endocrine, and nutrient factors all combine to affect the release of GLP-1. GLP-1 is secreted by the L-cells of the distal small intestine in a biphasic manner. The first phase occurs before nutrients make physical contact with the L-cells and lasts 10 to 15 minutes;

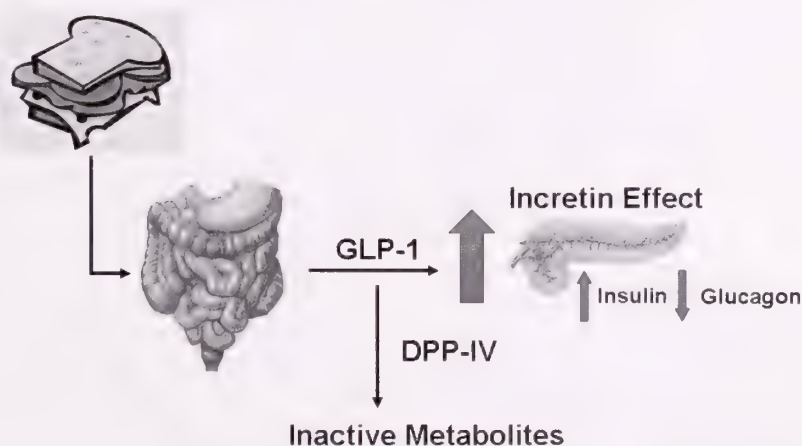


Fig 1. Food stimulates the gut to secrete GLP-1, which is responsible for the incretin effect. The DPP-IV enzyme rapidly breaks down this molecule into inactive metabolites. **Abbreviations:** DPP-IV, dipeptidyl peptidase IV; GLP-1, glucagon-like peptide 1. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com].

therefore, additional stimuli other than direct contact with the L-cells must be responsible for stimulating its release. GIP, which is released by the K-cells of the proximal small bowel, does not seem to have an effect on GLP-1 secretion in humans. Thus, GIP-mediated GLP-1 release does not occur.²⁴ However, gastrin-releasing peptide, which is produced by the postganglionic fibers of the vagus nerve that stimulate the G-cells of the stomach, does appear to have an effect on GLP-1 release. Experiments in rats that have undergone vagotomy have shown a disruption in fat-induced GLP-1 secretion, underscoring the importance of vagal nerve stimulation in the release of GLP-1.²⁵ Additionally, the importance of acetylcholine as a signaling molecule has been tested through the use of muscarinic receptor agonists to enhance GLP-1 release. Conversely, muscarinic receptor antagonists inhibit GLP-1 release in the presence of nutrients in the gut.^{22,23}

The 2-minute half-life of GLP-1 is the result of its degradation by the ubiquitous surface membrane enzyme dipeptidyl peptidase IV (DPP-IV), which cleaves dipeptides from proteins at the N-terminus when alanine or proline is in position number 2. The activity of DPP-IV is equal in the tissue of diabetics and nondiabetics, and this points toward a defect in the secretion of GLP-1 as the source of the deficiency in diabetics rather than an increase in breakdown by enzyme overactivity. Thus, inhibiting DPP-IV does not have antidiabetic effects as potent as GLP-1 administration (discussed later).²⁰

To enhance insulin release, GLP-1 binds to its receptor on the beta-cell and, via a series of intracellular events, increases insulin granule exocytosis.¹³ Additionally, GLP-1 has a direct effect on the transcriptional rate of the insulin gene, leading to increases in insulin stores.^{27,28} Thus, the overall effect on the beta-cell is to enhance insulin storage capability as well as release. As discussed later, GLP-1 also has antiapoptotic effects on the beta-cell in animal models, which can theoretically be exploited to prevent the progression of type 2 diabetes.

The biological effects of GLP-1 extend beyond its influence on the entero-insular axis. GLP-1 antagonizes gastric acid secretion and motility, which is normally stimulated by a meal. This delaying of gastric emptying may be of greater importance than the insulinotropic effect of GLP-1. Delayed gastric emptying itself leads to greater insulin release in response to a meal and creates a sensation of satiety.²⁹ GLP-1 binds to receptors in the paraventricular nucleus of the hypothalamus to directly suppress appetite via the central nervous system.³⁰ In addition, GLP-1 increases glucose disposal in muscle tissue by increasing glycogen

synthase activity and increasing glucose oxidative metabolism.¹³ GLP-1 also suppresses hepatic glucose production synergistically with insulin.³¹ Together, these functions further promote euglycemia (see Figure 2). Inhibiting GLP-1 breakdown with DPP-IV inhibitors affects some though not all of the target actions of the incretins, and this is discussed further later (see Figure 3).

CLINICAL TRIALS WITH GLUCAGON-LIKE PEPTIDE 1 RECEPTOR AGONISTS IN TYPE 2 DIABETES

Because of its short half-life, early studies of GLP-1 therapy were performed with continuous infusions to achieve a therapeutic effect. Weight loss (−1.4 kg), slower gastric emptying, and improved glycemic control [hemoglobin A1c (HbA1c) decreased by 1.3%] were demonstrated in patients with type 2 diabetes with continuous infusions of GLP-1.³²

The discovery of an incretin-like peptide (so-called incretin mimetics), isolated from the venom of *Heloderma suspectum* (the Gila monster), gave rise to more practical subcutaneous administration of a GLP-1 receptor agonist called exendin-4, the pharmacological equivalent of which is named exenatide.³³ Exenatide has a longer half-life than GLP-1 because it does not contain the amino acid sequences recognized by the DPP-IV enzyme and escapes degradation. Since its discovery, there have been a number of trials and significant clinical experience promoting its efficacy in controlling type 2 diabetes without causing weight gain (see Table 1).

Exenatide as an Add-On to Oral Therapy

Initial clinical trials with exenatide investigated its efficacy as add-on therapy with existing oral therapy when target glucose levels were not achieved. Adding exenatide to metformin or a sulfonylurea improved the homeostasis model assessment of insulin release (an index of beta-cell function) and lowered HbA1c in 11% of enrolled patients to the American Diabetes Association goal of less than 7% over the course of 1 month of therapy.³⁴ Three landmark 30-week AMIGO (AC2993 Diabetes Management for Improving Glucose Outcomes) trials further examined the efficacy of exenatide when added to oral therapies. When added to a sulfonylurea, exenatide demonstrated a dose-dependent decrease in HbA1C (−0.86% in the 10-μg twice daily group and −0.46% in the 5-μg twice daily group) as well

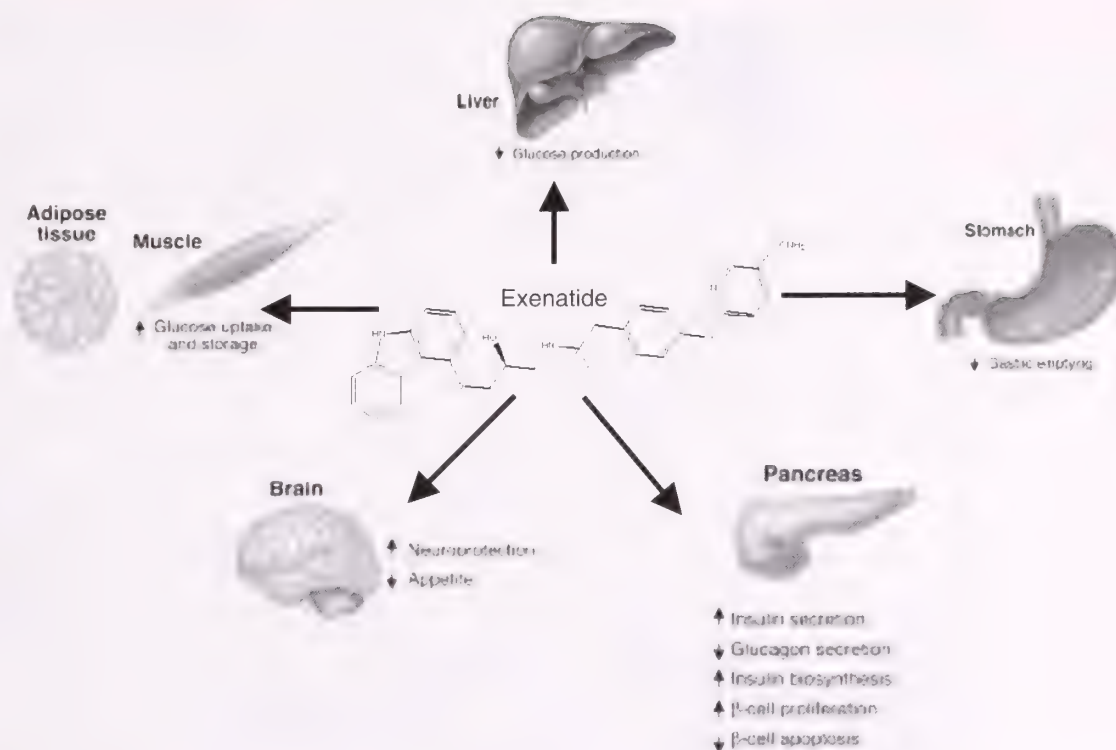


Fig 2. Tissue targets of the glucagon-like peptide 1 analogue exenatide. In addition to the incretin effect, glucagon-like peptide 1 delays gastric emptying, suppresses appetite in the central nervous system, increases glucose uptake in muscle and fat, and decreases glucose production in the liver. Adapted from *Gastroenterology*.¹³ [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

as a 1.6-kg weight loss in the higher dose group.³⁵ In the second trial, the addition of exenatide to metformin over a 30-week course led to a drop in HbA1c of -0.78 (from an average of 8.2%) as well as a dose-dependent loss of weight (-2.8 kg with $10\text{ }\mu\text{g}$ and -1.6 kg with $5\text{ }\mu\text{g}$).³⁶ In the third trial, exenatide was added to the regimens of patients on both metformin and a sulfonylurea with an HbA1c reduction of -1.0% in the $10\text{-}\mu\text{g}$ group and a weight decrease of -1.6 kg.³⁷

Exenatide also proved to be efficacious when added to TZDs. When exenatide was added to the regimens of patients on rosiglitazone or pioglitazone, a reduction in HbA1c of -0.98% and a reduction in body weight of -1.51 kg were seen in the therapy arm compared to the placebo arm. Forty percent of the treatment group and 15% of the placebo group experienced nausea, with a dropout rate of 16%.³⁸

The 3 AMIGO trials with exenatide^{35–37} were combined in a 2-year follow-up study assessing metabolic markers.³⁹ Of a total of 974 patients who received exenatide for 30 weeks, 283 subjects continued on the medication for 2 years. A reduction in HbA1c of -1.1% was sustained across the 2 years, with 50% of patients achieving HbA1c of $<7.0\%$.

Patients continued to lose weight on the drug, with an additional -2.2 kg lost over 2 years after the initial -2.1 kg lost in the 30-week trials. Measurements of insulin resistance, blood pressure, and liver function abnormalities all improved as well.³⁹ Once again, the most frequent adverse event was mild to moderate nausea. Patients on exenatide in the AMIGO trials have since been followed for 3 years with continued good results. Exenatide progressively reduced body weight from the baseline (-5.3 ± 0.4 kg at 3 years). The aforementioned reductions in HbA1c were sustained to 3 years ($-1.0\% \pm 0.1\%$), with 46% achieving HbA1c $<7\%$. Improvements in the lipid profile, blood pressure, and liver enzymes were also seen.⁴⁰

Exenatide Compared to Insulin Therapy

Next, investigators compared exenatide use versus insulin therapy in patients who had not achieved target glycemic control with oral agents alone. Potential benefits of adding exenatide instead of insulin to the regimen of patients failing on oral therapy would be promotion of weight loss instead of the usual weight gain seen with insulin therapy and lowered rates of hypoglycemia. Exenatide was studied in direct comparison to once daily insulin

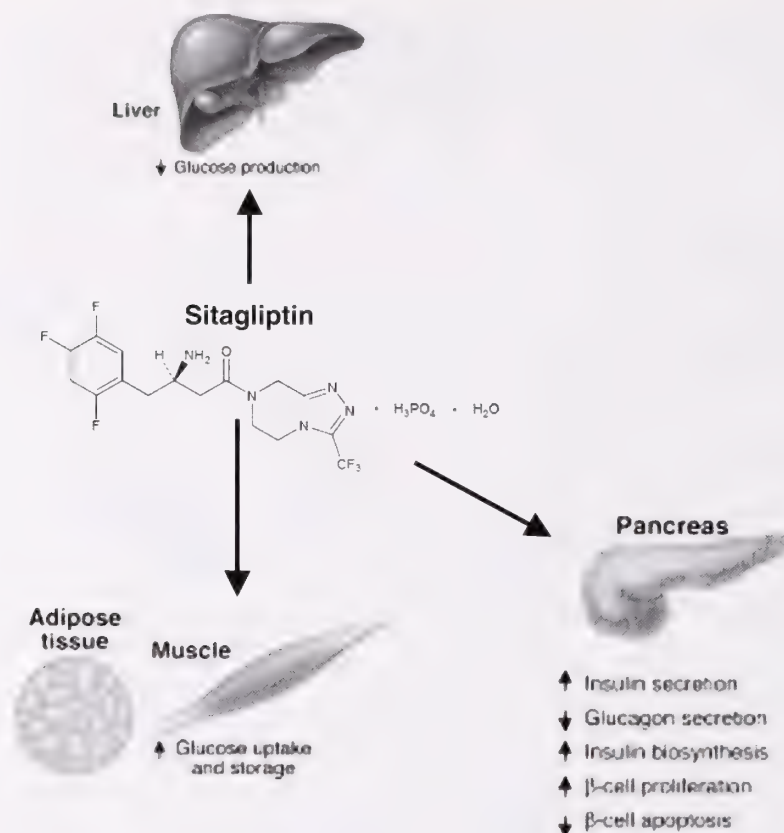


Fig 3. Dipeptidyl peptidase IV inhibitors raise glucagon-like peptide 1 to lower levels in comparison with exenatide administration. The effects are therefore limited to the liver, muscle, and pancreas. The effects on the stomach and brain are not seen. Adapted from *Gastroenterology*.¹³ [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

glargine in patients with suboptimally controlled diabetes on oral therapy.⁴¹ In both treatment arms of the study, the baseline HbA1C of 8.2% was lowered by -1.11% . Body weight decreased with exenatide by -2.3 kg and increased with glargine by 1.8 kg. As one might anticipate knowing the biology of the incretin effect, postprandial glucose excursions were better controlled in the exenatide group, whereas fasting sugars were better controlled in the glargine group. The prevalence of nausea was substantially higher in the exenatide group, with 57.1% of the patients experiencing this effect and 19.4% dropping out of the study.⁴¹

Exenatide has also been compared to twice daily insulin aspart, a short-acting insulin analogue, in addition to oral agents.⁴² Insulin aspart targets postprandial glucose levels like exenatide but unlike insulin glargine. In this trial, the withdrawal rate for exenatide was 21% versus 10% in the insulin group, with 33% of the exenatide group experiencing nausea. Both groups exhibited similar HbA1c-lowering effects, although the exenatide

group showed weight loss instead of weight gain (with an average weight difference of 5 kg between groups at the end of the study). The exenatide group also demonstrated less postprandial hyperglycemia.

In addition to assessing efficacy, researchers have also investigated the cost of exenatide compared to that of insulin glargine. A recent study in the United Kingdom conducted various cost-benefit analyses of using insulin glargine versus exenatide as an add-on to oral therapy. In every model tested, insulin glargine was more cost-effective in treating diabetes, even when weight gain and hypoglycemia were taken into account.⁴³ To address whether the nausea and gastrointestinal side effects of exenatide affected patient preference in choosing between insulin and exenatide add-on therapy, investigators conducted a quality-of-life study.⁴⁴ Patients who had been on either exenatide or insulin glargine added to oral therapy for 28 weeks were asked to fill out a questionnaire asking about side effects, drug tolerability, and satisfaction. Despite the increased gastrointestinal side effects of exenatide, patient

Table 1. Summary of Trials with Exenatide.

Reference	Medication and Dose	n	Study Weeks	Additional Medications	HbA1c Δ (%)	Weight Δ (kg)
Fineman <i>et al.</i> ³⁴	Exenatide, 0.08 μ g BID	109	4	SU $\pm$ metformin	-1.1	No data
	Exenatide, 0.08 μ g BID				-0.7	
	Exenatide, 0.08 μ g BID				-1.0	
	Placebo				-0.3	
Buse <i>et al.</i> ³⁹	Exenatide, 5 μ g BID	377	30	SU	-0.46	-0.9
	Exenatide, 10 μ g BID				-0.86	-1.6
	Placebo				+0.12	-0.6
DeFronzo <i>et al.</i> ³⁶	Exenatide, 5 μ g BID	336	30	Metformin	-0.4	-1.6
	Exenatide, 10 μ g BID				-0.78	-2.8
	Placebo				+0.08	-0.3
Kendall <i>et al.</i> ³⁷	Exenatide, 5 μ g BID	733	30	Metformin + SU	-0.6	-1.6
	Exenatide, 10 μ g BID				-0.8	-1.6
	Placebo				+0.2	-0.9
Zinman <i>et al.</i> ³⁸	Exenatide, 10 μ g BID	233	16	TZD + metformin	-0.9	-1.8
	Placebo				+0.1	-0.2
Heine <i>et al.</i> ⁴¹	Exenatide, 10 μ g BID	551	26	Metformin + SU	-1.1	-2.3
	Glargine				-1.1	+1.8
Nauck <i>et al.</i> ⁴²	Exenatide, 10 μ g BID	501	52	Metformin + SU	-1.0	-2.5
	Biphasic insulin				-0.9	+2.9
Kim <i>et al.</i> ⁴⁹	Exenatide LAR, 0.8 mg/week	45	15	Metformin or diet	-1.4	0.0
	Exenatide LAR, 2.0 mg/week				-1.7	-3.8
	Placebo				+0.4	0.0

NOTE: All administration was via subcutaneous injection.

Abbreviations: BID, 2 times a day; HbA1c, hemoglobin A1c; LAR, long-acting release; n, number of patients in the trial; SU, sulfonylurea; TZD, thiazolidinedione.

satisfaction was comparable to the satisfaction of those who used glargine.⁴⁴

Exenatide Use with Insulin

No randomized controlled trials have been conducted in humans to investigate the safety and efficacy of exenatide used in combination with insulin. However, this practice is common and has been studied retrospectively.⁴⁵ In 134 patients for whom exenatide was added to insulin therapy, a -0.87% average reduction of HbA1c was seen despite a 45% discontinuation of premeal insulin, a 9-unit reduction of average insulin use, and a decrease in the average number of insulin injections from 2 to 1. Exenatide use led to weight loss in 72% of patients, with an average loss of -5.2 kg. Exenatide was discontinued in 36% of patients because of gastrointestinal side effects, and the hypoglycemia rate in the treated group was 10%.⁴⁵

Long-Acting Glucagon-Like Peptide 1 Receptor Agonists

Further manipulations of GLP-1 analogues have led to a long-acting GLP-1 receptor agonist, liraglutide, which is a fatty acid derivative of GLP-1 that weakly binds to albumin; this results in a longer half-life.

This long-acting agonist can be administered once daily. An initial safety study of liraglutide showed daily doses achieving levels of medication similar to those achieved with twice daily doses of exenatide with a side-effect profile similar to that of exenatide.⁴⁶ An initial clinical trial in humans showed a positive drug effect over 1 week,⁴⁷ and a subsequent longer term trial also demonstrated HbA1c lowering when it was added to metformin, although 61% of patients receiving liraglutide experienced nausea versus 19.4% of the controls.⁴⁸

GLP-1 analogues that bind with more affinity to albumin are being developed for a once weekly dose, that is, exenatide long-acting release (LAR). Exenatide LAR forms microspheres in the bloodstream, which account for its long half-life. In a 15-week trial, once weekly exenatide LAR was added to metformin in patients with an average HbA1c of 8.5% and was dosed at 0.8 or 2.0 mg.⁴⁹ In comparison with placebo, HbA1C reductions of -1.4% and -1.8% at low and high doses, respectively, were achieved, but only those on the high dose lost weight (-3.8 kg). The longer acting molecule has a significant effect on lowering fasting glucose in comparison with the short-acting exenatide,⁵⁰ and this suggests that exenatide's negligible effect on fasting glucose may be the result of its short half-life. A reported side effect

was mild nausea, and unlike previous trials, there was no dropout in the exenatide LAR group. In a 30-week noninferiority study comparing twice daily exenatide to once weekly exenatide LAR, 295 patients with an average HbA1c of 8.3% were randomized to either intervention. A greater proportion of the weekly dosed patients (77%) reached HbA1c levels <7.0 in comparison with the twice daily dosed patients (61%; $P = 0.0039$).⁵¹

Dipeptidyl Peptidase IV Inhibitors

By preventing the breakdown of GLP-1, the new class of DPP-IV inhibitors offers a different method to enhance the incretin effect in the treatment of type 2 diabetes (Table 2). There is currently only 1 US Food and Drug Administration (FDA)-approved DPP-IV inhibitor available: sitagliptin. Vildagliptin, saxagliptin, and alogliptin are not yet FDA-approved. Initial studies looking at the effects of inhibiting this enzyme revealed effects similar to those from using GLP-1 analogue-based therapies: decreased glucagon levels, increased postprandial insulin levels, and decreased postprandial glucose levels.⁵² These agents also have demonstrated an improvement in beta-cell function as measured by mathematical

models that take into account glucose and insulin levels before and after treatment.⁵³

Clinical studies looking at the efficacy and safety of adding sitagliptin to metformin monotherapy^{54–56} and to pioglitazone (a TZD)⁵⁷ have yielded positive results, with HbA1c lowering of an additional 0.7% to 1.1% from a baseline HbA1c of approximately 8.0%. Adding sitagliptin to a sulfonylurea seems less effective, with HbA1c reductions of 0.3% to 0.6%.^{58,59} Sitagliptin use as monotherapy has also been studied, with a reduction in HbA1c of –0.6% from a baseline of 8.0%.⁶⁰ Adding a DPP-IV inhibitor to insulin therapy does not seem beneficial, as the HbA1c-lowering effect was only 0.2% over placebo when sitagliptin was added to insulin therapy.⁶¹

Head-to-head trials of DPP-IV inhibitors versus established diabetes medications have revealed the relative inferiority of the drugs compared to their predecessors, although they have fewer adverse side effects such as hypoglycemia and weight gain. When metformin and vildagliptin were directly compared as first-line therapy in drug-naïve patients with type 2 diabetes, the metformin group achieved a 0.4% lower HbA1c than the vildagliptin group after 52 weeks.⁶² When the sulfonylurea glipizide or sitagliptin was added to the regimen of patients failing on monotherapy with metformin, both groups

Table 2. Summary of Trials with Dipeptidyl Peptidase IV Inhibitors.

Reference	Medication and Dose	n	Study Weeks	Additional Medication	HbA1c Δ (%)	Weight Δ (kg)
Fonseca <i>et al.</i> ⁶¹	Vildagliptin, 50 mg BID	152	24	Insulin	–0.5	+1.3
	Placebo				–0.2	+0.6
Charbonnel <i>et al.</i> ⁵⁴	Sitagliptin, 100 mg QD	701	24	Metformin	–0.67	–0.6
	Placebo				–0.02	–0.6
Goldstein <i>et al.</i> ⁵⁵	Sitagliptin, 100 mg/metformin, 1 g	1091	24	None	–1.57	–0.6 to –1.9
	Sitagliptin, 100 mg/metformin, 2 g				–2.07	
	Metformin, 1 g QD				–0.99	0.0
	Metformin, 1 g BID				–1.30	
	Sitagliptin, 100 mg QD				–0.83	
Rosenstock <i>et al.</i> ⁵⁷	Sitagliptin, 100 mg PO QD	353	24	Pioglitazone	–0.80	+1.8
	Placebo				–0.10	+1.5
Garber <i>et al.</i> ⁵⁸	Sitagliptin, 50 mg PO QD	515	24	SU	–0.58	–0.1
	Sitagliptin, 50 mg PO BID				–0.63	+1.3
	Placebo				+0.07	–0.4
Raz <i>et al.</i> ⁵⁶	Sitagliptin, 100 mg PO QD	190	30	Metformin	–1.0	–0.5
	Placebo				0.0	–0.5
Hermansen <i>et al.</i> ⁵⁹	Sitagliptin, 100 mg PO QD	441	24	SU ± metformin	–0.45	+0.8
	Placebo				+0.28	–0.4
Schweizer <i>et al.</i> ⁶²	Vildagliptin, 100 mg PO QD	526	52	None	–1.0	+0.3
	Metformin, 1 g PO BID				–1.9	–1.9
Pan <i>et al.</i> ⁶³	Vildagliptin, 50 mg PO BID	661	24	None	–1.4	–0.4
	Acarbose, 100 mg PO TID				–1.3	–1.7

Abbreviations: BID, 2 times a day; HbA1c, hemoglobin A1c; n, number of patients in the trial; PO, by mouth; QD, daily; SU, sulfonylurea; TID, 3 times a day.

had an HbA1c reduction of -0.7 ; however, the dose of glipizide used was not maximal (10 mg). If maximum doses of both drugs had been used, glipizide may have shown superiority. The rates of hypoglycemia were markedly lower in the sitagliptin group compared to the glipizide group (32% versus 5%), and a -1.5 -kg weight loss was seen in the sitagliptin group versus a weight gain of 1.1 kg in the glipizide group.⁴² Pioglitazone and vildagliptin had similar HbA1c-lowering effects in a head-to-head trial of around -1.0% , with a weight-neutral effect seen with vildagliptin and a 1.6-kg weight gain seen with pioglitazone. Compared to the alpha-glucosidase inhibitor acarbose, vildagliptin lowered HbA1c from a level of 8.6% by a similar amount (-1.4% versus -1.3%), with 25% of patients experiencing gastrointestinal side effects in the acarbose group versus 12% in the vildagliptin arm.⁶³ Patients taking acarbose experienced a -1.7 -kg weight loss versus -0.4 kg in the vildagliptin group.⁶³

When directly compared, the effects of exenatide seem more potent than the effects of the DPP-IV inhibitors.⁶⁴ In a short 1-month study, patients treated with metformin were given either exenatide or sitagliptin and after 2 weeks were switched to the other medication. Measurements of glucoregulatory hormones and glucose levels were recorded while they were on the 2 medications. Average 2-hour postprandial glucoses were significantly lower in the exenatide group compared to the sitagliptin group (133 versus 208 mg/dL), whereas fasting glucose levels were similar in the 2 groups. Exenatide had a more potent insulinogenic effect than sitagliptin and was more effective at lowering postprandial glucagon levels. Overall, this study illustrates the higher potency of the exendin-4 molecule, which is not degraded by DPP-IV, over the inhibition of this enzyme for boosting endogenous levels.

ADVERSE EFFECTS OF INCRETIN THERAPY

Some clinicians have expressed concern over the risk of pancreatitis associated with the use of exenatide. There have been 30 incidents of pancreatitis with exenatide use, all of which had 1 of the following risk factors for pancreatitis: gallstones, alcohol abuse, or marked elevations of triglycerides.⁶⁵ Six cases of hemorrhagic pancreatitis have been reported to the FDA with 2 deaths, and exenatide was stopped in all patients, although a direct causal relationship remains unclear.⁶⁶ Exenatide is contraindicated in patients with gastroparesis, as further delay of gastric

emptying could predispose patients to intestinal perforation, bezoar formation, or pancreatitis.⁶⁵ Overall, the medication has been shown to be safe, but it should be used with caution in those with risk factors for pancreatitis.

DPP-IV is a ubiquitous enzyme on the cell membrane and is present in T-cell lymphocyte; this raises concern among some clinicians about possible immune-suppressive effects of this drug. A meta-analysis found that patients taking DPP-IV inhibitors were 1.2 times as likely to develop nasopharyngitis and 1.5 times as likely to develop a urinary tract infection (UTI).⁶⁷ The studies with sitagliptin were 10 times as large as the studies with vildagliptin, which is not FDA-approved. When we look at sitagliptin studies only by taking out the vildagliptin studies, we find that the risk of UTI was blunted (3.1% in treated patients and 2.6% in controls). Because the larger studies showed less risk, it is possible that as more data become available on adverse events, these risks may no longer be statistically significant. The absolute percentages in this analysis of patients with nasopharyngitis were 6.4% in the sitagliptin group and 6.1% in the controls. The slight increase over baseline nasopharyngitis was statistically significant but likely not clinically significant. A proposed mechanism explaining the slight increase is as follows: DPP-IV does inactivate the proinflammatory peptide released by sensory nerve fibers in the nasal mucosa, and this could explain the increased risk and may mean that caution should be used in patients with chronic sinusitis.⁶⁸ The mechanism by which patients may have an increased risk specifically for UTI is undetermined.

POTENTIAL USES OF INCRETIN-BASED THERAPY

Investigators have looked at a possible role of GLP-1-based therapies in treating polycystic ovary syndrome, a syndrome characterized by increased androgen levels, insulin resistance, and anovulation with irregular menstrual cycles.⁶⁹ Hyperinsulinemia stimulates excess androgen production by the ovary, so with a reduction of insulin resistance, ovulation can be restored. Two other agents used to treat insulin resistance, metformin and TZDs, have been shown to decrease androgen levels and increase ovulation in women with polycystic ovary syndrome.^{70,71} A 24-week study examining the use of exenatide, metformin, or both in restoring ovulation, decreasing androgen levels, and decreasing insulin resistance showed the best results with combination therapy.⁶⁹

Metformin alone restored ovulation in 30% of patients, whereas the combination of metformin and exenatide restored ovulation in 85% of patients ($P < 0.01$). In the study, 30% of the patients had impaired glucose tolerance (IGT), and 65% of those women no longer had IGT by the end of the study. The decrease in insulin resistance was related to the weight loss seen in these women treated with both agents.

GLP-1 has been shown to have antiapoptotic effects and in effect increases beta-cell mass in animal models (see Figure 4).^{72,73} These studies point toward the potential role of incretin-based therapy in treating prediabetes or IGT. Vildagliptin was studied in patients with IGT with positive results.⁷⁴ After 12 weeks on therapy, patients on vildagliptin, compared to patients on placebo, had improved beta-cell function by measures of insulin-to-glucose ratios and improved alpha-cell function as measured by decreased glucagon-to-glucose ratios postprandially.

The trial did not show an effect on the homeostasis model assessment of insulin release, likely because the medication had no effect on fasting glucose levels.

In addition to a potential role in treating prediabetes, incretin-based therapies may have a role in weight loss in obese patients. Studies looking at this effect are currently ongoing, and those results, if positive, could lead to novel medical interventions for weight loss.

OTHER GUT HORMONES

In addition to GLP-1 and GIP, other small peptide molecules produced by the gastrointestinal tract and pancreas have been discovered in recent years that act on different areas of the brain (see Figure 5) and on the gastrointestinal tract itself to regulate food intake. One of these molecules, amylin, is

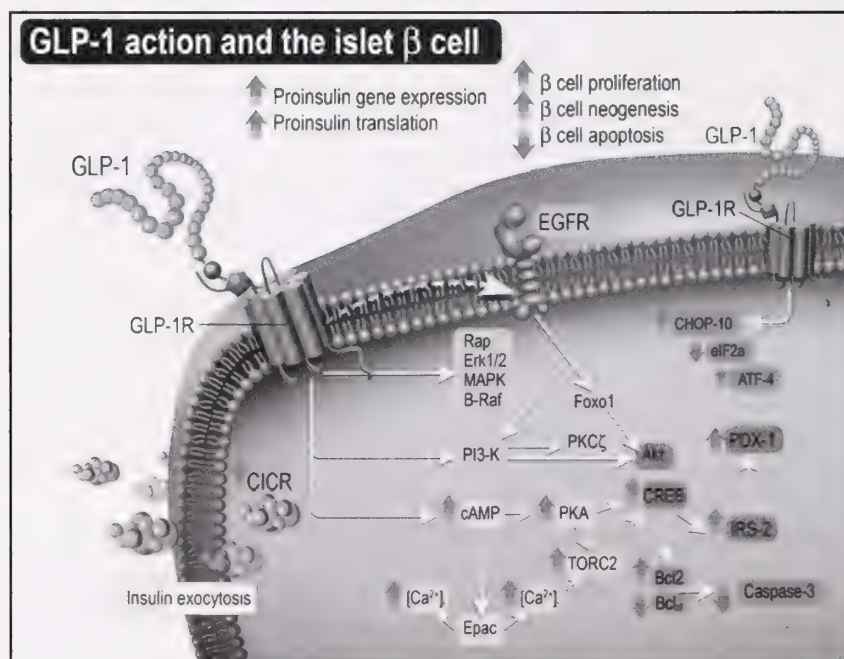


Fig 4. The animal model of GLP-1 stimulation in the beta-cell of the pancreas. Intracellular signaling pathways are activated that promote proliferation and decrease apoptosis. **Abbreviations:** Akt, protein kinase B; ATF-4, activating transcription factor 4; B-Raf, v-raf murine sarcoma viral oncogene homolog B1; Bcl, basal cell lymphoma; cAMP, cyclic adenosine monophosphate; CHOP-10, C/EBP homologous protein 10; CICR, calcium-induced calcium release; CREB, cyclic adenosine monophosphate response element binding; EGFR, epidermal growth factor receptor; eIF2a, eukaryotic translation initiation factor 2a; Epac, exchange protein directly activated by cyclic adenosine monophosphate; Erk1/2, extracellular signal-regulated kinase 1/2; Foxo1, forkhead box O1; GLP, glucagon-like peptide; IRS-2, insulin receptor substrate 2; MAPK, mitogen-activated protein kinase; PDX-1, pancreatic and duodenal homeobox 1; PI3-K, phosphoinositide 3-kinase; PKA, protein kinase A; PKCζ, protein kinase C zeta; Rap, repressor activator protein; TORC2, TOR complex 2. Reprinted with permission from *Cell Metabolism*.⁷⁹ Copyright 2006, Elsevier, Inc. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

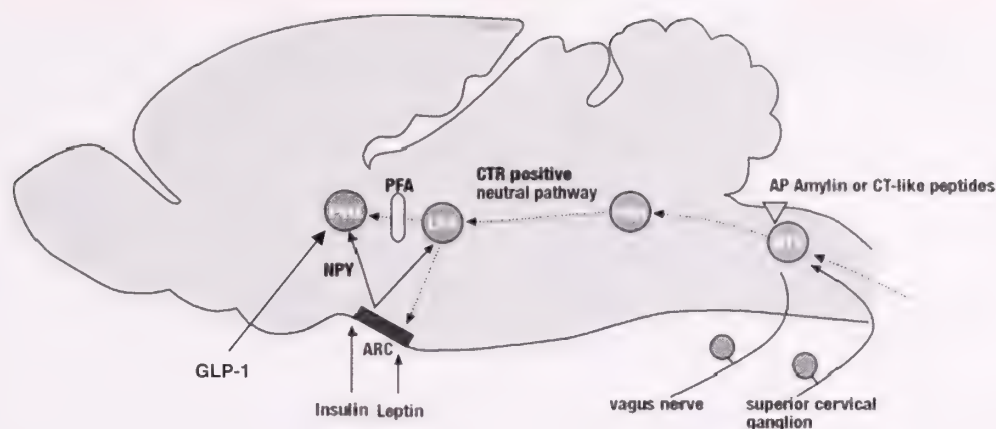


Fig 5. Actions of hormones on appetite regulation centers in the brain. A solid line indicates up-regulation, and a dashed line indicates inhibition. **Abbreviations:** AP, area postrema; ARC, arcuate nucleus; CT, calcitonin; CTR, calcitonin receptor; GLP-1, glucagon-like peptide 1; LHA, lateral hypothalamic area; NPY, neuropeptide Y; NTS, nucleus tractus solitarius; PBN, parabrachial nuclear complex; PFA, paraventricular nucleus. Adapted from *Science World Journal*.⁸⁰ [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 3. *Incretin Mimetics Available for the Treatment of Type 2 Diabetes Mellitus.*

Medication	Formulation	Usual Daily Dose	Monthly Cost
Sitagliptin (Januvia)	25-, 50-, and 100-mg tabs	100 mg PO once	\$180.60
Metformin/sitagliptin (Janumet)	500/50- and 1000/50-mg tabs	500/50 mg PO BID	\$181.80
Exenatide (Byetta)	1.2- and 2.5-mL prefilled syringes (250 µg/mL)	10 µg SC BID before meals	\$213.03
Pramlintide (Amylin)	5-mL vial (0.6 mg/mL) and 1.5- and 2.7-mL prefilled pens (1000 µg/mL)	60–120 µg SC TID before meals	\$231.60

Abbreviations: BID, 2 times a day; PO, by mouth; SC, subcutaneously; TID, 3 times a day.

secreted by the beta-cells of the pancreas in amounts equimolar to insulin. Amylin promotes satiety and delays gastric emptying synergistically with GLP-1.⁷⁵ Amylin is deficient in patients with type 1 diabetes and in patients with type 2 diabetes, and its synthetic analogue pramlintide has been investigated and approved by the FDA for the treatment of both of these conditions.⁷⁶ In a year-long clinical trial, pramlintide was added to insulin therapy in patients with type 2 diabetes as premeal injections, and it lowered HbA1c by an average of -0.6% in comparison with placebo, with mild nausea as the most common side effect and no increase in rates of hypoglycemia.⁷⁷ Similarly to exenatide, use of pramlintide promotes weight loss. In a year-long study in patients with type 2 diabetes, the addition of twice daily high-dose (360-µg) pramlintide led to a sustained average weight loss of -3.2 kg in comparison with placebo, with mild to moderate nausea occurring in 29% of patients.⁷⁸ Because

pramlintide is FDA-approved for use with insulin therapy but exenatide currently is not, pramlintide should be considered for a suboptimally controlled patient on insulin for both improved postprandial glycemic control and weight loss. It should be noted that amylin is not FDA-approved for use without insulin.

CONCLUSION

Manipulations of the entero-insular axis by targeting the incretin effect of GLP-1 have proven to be safe and efficacious in treating type 2 diabetes (Table 3). Importantly, these agents are weight-neutral (the DPP-IV inhibitors) or promote weight loss (GLP-1 receptor agonists). By preventing weight gain, these drugs treat type 2 diabetes without exacerbating the underlying inflammatory pathophysiology found

in obesity that contributes to insulin resistance and beta-cell dysfunction. Mounting in vitro data demonstrate that these medications preserve beta-cell function, may prevent the transition from prediabetes to diabetes, and slow the progression of type 2 diabetes.

Though perhaps not justified to be used as first-line therapy across the board given the inferiority of the DPP-IV inhibitors versus other oral hypoglycemic drugs and the cost of exenatide compared to that of other agents, these medications should play an important role in managing type 2 diabetes. For example, patients with renal failure as a contraindication to metformin can be safely and effectively treated with sitagliptin at a reduced dose. Alternatively, patients failing on multiple oral agents may benefit from the addition of exenatide instead of insulin to help promote weight loss in addition to glycemic control (if c-peptide levels are present). There is promise that the incretins may prevent or delay the progression of many patients with type 2 diabetes to insulin therapy by virtue of their ability to prevent beta-cell apoptosis and promote beta-cell neogenesis in animal models. In addition, subhuman primate studies and ultimately human studies will be necessary to demonstrate that the animal data can be extrapolated to the clinic. These further studies will be necessary to provide us with the full therapeutic repertoire of the incretins.

DISCLOSURES

Potential conflict of interest: Dr Rayfield participates in the speaker's bureaus for Sanofi Aventis, Amylin, Eli Lilly; Abbott; Merck; and Novartis.

REFERENCES

1. Mokdad AH, Ford ES, Bowman BA, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *JAMA* 2003; 289: 76–79.
2. Anderson JW, Kendall CW, Jenkins DJ. Importance of weight management in type 2 diabetes: review with meta-analysis of clinical studies. *J Am Coll Nutr* 2003; 22: 331–339.
3. Bray GA, Jablonski KA, Fujimoto WY, et al. Relation of central adiposity and body mass index to the development of diabetes in the Diabetes Prevention Program. *Am J Clin Nutr* 2008; 87: 1212–1218.
4. Balkau B, Deanfield JE, Despres JP, et al. International Day for the Evaluation of Abdominal Obesity (IDEA): a study of waist circumference, cardiovascular disease, and diabetes mellitus in 168,000 primary care patients in 63 countries. *Circulation* 2007; 116: 1942–1951.
5. Heidemann C, Sun Q, van Dam RM, et al. Total and high-molecular-weight adiponectin and resistin in relation to the risk for type 2 diabetes in women. *Ann Intern Med* 2008; 149: 307–316.
6. Pilz S, Maerz W, Weihrauch G, et al. Adiponectin serum concentrations in men with coronary artery disease: the Ludwigshafen Risk and Cardiovascular Health (LURIC) study. *Clin Chim Acta* 2006; 364: 251–255.
7. de Souza CJ, Eckhardt M, Gagen K, et al. Effects of pioglitazone on adipose tissue remodeling within the setting of obesity and insulin resistance. *Diabetes* 2001; 50: 1863–1871.
8. Hirose H, Kawai T, Yamamoto Y, et al. Effects of pioglitazone on metabolic parameters, body fat distribution, and serum adiponectin levels in Japanese male patients with type 2 diabetes. *Metabolism* 2002; 51: 314–317.
9. Nakamura T, Funahashi T, Yamashita S, et al. Thiazolidinedione derivative improves fat distribution and multiple risk factors in subjects with visceral fat accumulation—double-blind placebo-controlled trial. *Diabetes Res Clin Pract* 2001; 54: 181–190.
10. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002; 346: 393–403.
11. Van de Laar FA, Lucassen PL, Akkermans RP, et al. Alpha-glucosidase inhibitors for people with impaired glucose tolerance or impaired fasting blood glucose. *Cochrane Database Syst Rev* 2006; CD005061.
12. van de Laar FA, Lucassen PL, Akkermans RP, et al. Alpha-glucosidase inhibitors for patients with type 2 diabetes: results from a Cochrane systematic review and meta-analysis. *Diabetes Care* 2005; 28: 154–163.
13. Baggio LL, Drucker DJ. Biology of incretins: GLP-1 and GIP. *Gastroenterology* 2007; 132: 2131–2157.
14. Perley MJ, Kipnis DM. Plasma insulin responses to oral and intravenous glucose: studies in normal and diabetic subjects. *J Clin Invest* 1967; 46: 1954–1962.
15. Elrick H, Stimmler L, Hlad CJ Jr, Arai Y. Plasma insulin response to oral and intravenous glucose administration. *J Clin Endocrinol Metab* 1964; 24: 1076–1082.
16. Nauck MA, Homberger E, Siegel EG, et al. Incretin effects of increasing glucose loads in man calculated from venous insulin and C-peptide responses. *J Clin Endocrinol Metab* 1986; 63: 492–498.
17. Gerich JE. Metabolic abnormalities in impaired glucose tolerance. *Metabolism* 1997; 46: 40–43.
18. Vilsboll T, Krarup T, Dæacon CF, et al. Reduced postprandial concentrations of intact biologically active glucagon-like peptide 1 in type 2 diabetic patients. *Diabetes* 2001; 50: 609–613.
19. Vilsboll T, Krarup T, Madsbad S, Holst JJ. Both GLP-1 and GIP are insulinotropic at basal and postprandial glucose levels and contribute nearly equally to the incretin effect of a meal in healthy subjects. *Regul Pept* 2003; 114: 115–121.
20. Quidussi S, Vahl TP, Hanson K, et al. Differential effects of acute and extended infusions of glucagon-like peptide-1 on first- and second-phase insulin secretion in diabetic and nondiabetic humans. *Diabetes Care* 2003; 26: 791–798.
21. Holst JJ, Gromada J. Role of incretin hormones in the regulation of insulin secretion in diabetic and nondiabetic humans. *Am J Physiol Endocrinol Metab* 2004; 287: E199–E206.

22. Anini Y, Brubaker PL. Muscarinic receptors control glucagon-like peptide 1 secretion by human endocrine L cells. *Endocrinology* 2003; 144: 3244–3250.
23. Anini Y, Hansotia T, Brubaker PL. Muscarinic receptors control postprandial release of glucagon-like peptide-1: in vivo and in vitro studies in rats. *Endocrinology* 2002; 143: 2420–2426.
24. Reimann F, Gribble FM. Glucose-sensing in glucagon-like peptide-1-secreting cells. *Diabetes* 2002; 51: 2757–2763.
25. Rocca AS, Brubaker PL. Role of the vagus nerve in mediating proximal nutrient-induced glucagon-like peptide-1 secretion. *Endocrinology* 1999; 140: 1687–1694.
26. Deacon CF, Nauck MA, Toft-Nielsen M, et al. Both subcutaneously and intravenously administered glucagon-like peptide I are rapidly degraded from the NH₂-terminus in type II diabetic patients and in healthy subjects. *Diabetes* 1995; 44: 1126–1131.
27. Alarcon C, Wicksteed B, Rhodes CJ. Exendin 4 controls insulin production in rat islet beta cells predominantly by potentiation of glucose-stimulated proinsulin biosynthesis at the translational level. *Diabetologia* 2006; 49: 2920–2929.
28. Fehmann HC, Habener JF. Insulinotropic hormone glucagon-like peptide-I(7–37) stimulation of proinsulin gene expression and proinsulin biosynthesis in insulinoma beta TC-1 cells. *Endocrinology* 1992; 130: 159–166.
29. Meier JJ, Kemmeries G, Holst JJ, Nauck MA. Erythromycin antagonizes the deceleration of gastric emptying by glucagon-like peptide 1 and unmasks its insulinotropic effect in healthy subjects. *Diabetes* 2005; 54: 2212–2218.
30. Turton MD, O'Shea D, Gunn I, et al. A role for glucagon-like peptide-1 in the central regulation of feeding. *Nature* 1996; 379: 69–72.
31. Prigeon RL, Quddusi S, Paty B, D'Alessio DA. Suppression of glucose production by GLP-1 independent of islet hormones: a novel extrapancreatic effect. *Am J Physiol Endocrinol Metab* 2003; 285: E701–E707.
32. Zander M, Madsbad S, Madsen JL, Holst JJ. Effect of 6-week course of glucagon-like peptide 1 on glycaemic control, insulin sensitivity, and beta-cell function in type 2 diabetes: a parallel-group study. *Lancet* 2002; 359: 824–830.
33. Eng J, Kleinman WA, Singh L, et al. Isolation and characterization of exendin-4, an exendin-3 analogue, from *Heloderma suspectum* venom. Further evidence for an exendin receptor on dispersed acini from guinea pig pancreas. *J Biol Chem* 1992; 267: 7402–7405.
34. Fineman MS, Bicsak TA, Shen LZ, et al. Effect on glycemic control of exenatide (synthetic exendin-4) additive to existing metformin and/or sulfonylurea treatment in patients with type 2 diabetes. *Diabetes Care* 2003; 26: 2370–2377.
35. Buse JB, Henry RR, Han J, et al. Effects of exenatide (exendin-4) on glycemic control over 30 weeks in sulfonylurea-treated patients with type 2 diabetes. *Diabetes Care* 2004; 27: 2628–2635.
36. DeFronzo RA, Ratner RE, Han J, et al. Effects of exenatide (exendin-4) on glycemic control and weight over 30 weeks in metformin-treated patients with type 2 diabetes. *Diabetes Care* 2005; 28: 1092–1100.
37. Kendall DM, Riddle MC, Rosenstock J, et al. Effects of exenatide (exendin-4) on glycemic control over 30 weeks in patients with type 2 diabetes treated with metformin and a sulfonylurea. *Diabetes Care* 2005; 28: 1083–1091.
38. Zinman B, Hoogwerf BJ, Duran Garcia S, et al. The effect of adding exenatide to a thiazolidinedione in suboptimally controlled type 2 diabetes: a randomized trial. *Ann Intern Med* 2007; 146: 477–485.
39. Buse JB, Klonoff DC, Nielsen LL, et al. Metabolic effects of two years of exenatide treatment on diabetes, obesity, and hepatic biomarkers in patients with type 2 diabetes: an interim analysis of data from the open-label, uncontrolled extension of three double-blind, placebo-controlled trials. *Clin Ther* 2007; 29: 139–153.
40. Klonoff DC, Buse JB, Nielsen LL, et al. Exenatide effects on diabetes, obesity, cardiovascular risk factors and hepatic biomarkers in patients with type 2 diabetes treated for at least 3 years. *Curr Med Res Opin* 2008; 24: 275–286.
41. Heine RJ, Van Gaal LF, Johns D, et al. Exenatide versus insulin glargine in patients with suboptimally controlled type 2 diabetes: a randomized trial. *Ann Intern Med* 2005; 143: 559–569.
42. Nauck MA, Duran S, Kim D, et al. A comparison of twice-daily exenatide and biphasic insulin aspart in patients with type 2 diabetes who were suboptimally controlled with sulfonylurea and metformin: a non-inferiority study. *Diabetologia* 2007; 50: 259–267.
43. Woehl A, Evans M, Tetlow AP, McEwan P. Evaluation of the cost effectiveness of exenatide versus insulin glargine in patients with sub-optimally controlled type 2 diabetes in the United Kingdom. *Cardiovasc Diabetol* 2008; 7: 24.
44. Secnik Boye K, Matza LS, Oglesby A, et al. Patient-reported outcomes in a trial of exenatide and insulin glargine for the treatment of type 2 diabetes. *Health Qual Life Outcomes* 2006; 4: 80.
45. Sheffield CA, Kane MP, Busch RS, et al. Safety and efficacy of exenatide in combination with insulin in patients with type 2 diabetes mellitus. *Endocr Pract* 2008; 14: 285–292.
46. Agerso H, Jensen LB, Elbrond B, et al. The pharmacokinetics, pharmacodynamics, safety and tolerability of NN2211, a new long-acting GLP-1 derivative, in healthy men. *Diabetologia* 2002; 45: 195–202.
47. Degn KB, Juhl CB, Sturis J, et al. One week's treatment with the long-acting glucagon-like peptide 1 derivative liraglutide (NN2211) markedly improves 24-h glycemia and alpha- and beta-cell function and reduces endogenous glucose release in patients with type 2 diabetes. *Diabetes* 2004; 53: 1187–1194.
48. Nauck MA, Hompesch M, Filipczak R, et al. Five weeks of treatment with the GLP-1 analogue liraglutide improves glycaemic control and lowers body weight in subjects with type 2 diabetes. *Exp Clin Endocrinol Diabetes* 2006; 114: 417–423.
49. Kim D, MacConell L, Zhuang D, et al. Effects of once-weekly dosing of a long-acting release formulation of exenatide on glucose control and body weight in subjects with type 2 diabetes. *Diabetes Care* 2007; 30: 1487–1493.
50. Juhl CB, Hollingdal M, Sturis J, et al. Bedtime administration of NN2211, a long-acting GLP-1 derivative, substantially reduces fasting and

- postprandial glycemia in type 2 diabetes. *Diabetes* 2002; 51: 121–129.
51. Drucker DJ, Buse JB, Taylor K, et al. Exenatide once weekly versus twice daily for the treatment of type 2 diabetes: a randomised, open-label, non-inferiority study. *Lancet* 2008; 372: 1240–1250.
 52. Ahren B, Landin-Olsson M, Jansson PA, et al. Inhibition of dipeptidyl peptidase-4 reduces glycemia, sustains insulin levels, and reduces glucagon levels in type 2 diabetes. *J Clin Endocrinol Metab* 2004; 89: 2078–2084.
 53. Mari A, Sallas WM, He YL, et al. Vildagliptin, a dipeptidyl peptidase-IV inhibitor, improves model-assessed beta-cell function in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2005; 90: 4888–4894.
 54. Charbonnel B, Karasik A, Liu J, et al. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor sitagliptin added to ongoing metformin therapy in patients with type 2 diabetes inadequately controlled with metformin alone. *Diabetes Care* 2006; 29: 2638–2643.
 55. Goldstein BJ, Feinglos MN, Luncsford JK, et al. Effect of initial combination therapy with sitagliptin, a dipeptidyl peptidase-4 inhibitor, and metformin on glycemic control in patients with type 2 diabetes. *Diabetes Care* 2007; 30: 1979–1987.
 56. Raz I, Chen Y, Wu M, et al. Efficacy and safety of sitagliptin added to ongoing metformin therapy in patients with type 2 diabetes. *Curr Med Res Opin* 2008; 24: 537–550.
 57. Rosenstock J, Brazg R, Andryuk PJ, et al. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor sitagliptin added to ongoing pioglitazone therapy in patients with type 2 diabetes: a 24-week, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Clin Ther* 2006; 28: 1556–1568.
 58. Garber AJ, Foley JE, Banerji MA, et al. Effects of vildagliptin on glucose control in patients with type 2 diabetes inadequately controlled with a sulphonylurea. *Diabetes Obes Metab* 2008; 10: 1047–1056.
 59. Hermansen K, Kipnes M, Luo E, et al. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor, sitagliptin, in patients with type 2 diabetes mellitus inadequately controlled on glimepiride alone or on glimepiride and metformin. *Diabetes Obes Metab* 2007; 9: 733–745.
 60. Raz I, Hanefeld M, Xu L, et al. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy in patients with type 2 diabetes mellitus. *Diabetologia* 2006; 49: 2564–2571.
 61. Fonseca V, Schweizer A, Albrecht D, et al. Addition of vildagliptin to insulin improves glycaemic control in type 2 diabetes. *Diabetologia* 2007; 50: 1148–1155.
 62. Schweizer A, Couturier A, Foley JE, Dejager S. Comparison between vildagliptin and metformin to sustain reductions in HbA(1c) over 1 year in drug-naïve patients with type 2 diabetes. *Diabet Med* 2007; 24: 955–961.
 63. Pan C, Yang W, Barona JP, et al. Comparison of vildagliptin and acarbose monotherapy in patients with type 2 diabetes: a 24-week, double-blind, randomized trial. *Diabet Med* 2008; 25: 435–441.
 64. Defronzo RA, Okerson T, Viswanathan P, et al. Effects of exenatide versus sitagliptin on postprandial glucose, insulin and glucagon secretion, gastric emptying, and caloric intake: a randomized, cross-over study. *Curr Med Res Opin* 2008; 24: 2943–2952.
 65. Ahmad SR, Swann J. Exenatide and rare adverse events. *N Engl J Med* 2008; 358: 1970–1971; discussion 1971–1972.
 66. Kuehn BM. Exenatide risk. *JAMA* 2008; 12: 1403.
 67. Amori RE, Lau J, Pittas AG. Efficacy and safety of incretin therapy in type 2 diabetes: systematic review and meta-analysis. *JAMA* 2007; 298: 194–206.
 68. Grouzmann E, Monod M, Landis BN, Lacroix JS. Adverse effects of incretin therapy for type 2 diabetes. *JAMA* 2007; 298: 1759–1760; author reply 1760.
 69. Elkind-Hirsch K, Marrioneaux O, Bhushan M, et al. Comparison of single and combined treatment with exenatide and metformin on menstrual cyclicity in overweight women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2008; 93: 2670–2678.
 70. Azziz R, Ehrmann D, Legro RS, et al. Troglitazone improves ovulation and hirsutism in the polycystic ovary syndrome: a multicenter, double blind, placebo-controlled trial. *J Clin Endocrinol Metab* 2001; 86: 1626–1632.
 71. Velazquez EM, Mendoza S, Hamer T, et al. Metformin therapy in polycystic ovary syndrome reduces hyperinsulinemia, insulin resistance, hyperandrogenemia, and systolic blood pressure, while facilitating normal menses and pregnancy. *Metabolism* 1994; 43: 647–654.
 72. Drucker DJ. Glucagon-like peptide-1 and the islet beta-cell: augmentation of cell proliferation and inhibition of apoptosis. *Endocrinology* 2003; 144: 5145–5148.
 73. Drucker DJ. Glucagon-like peptides: regulators of cell proliferation, differentiation, and apoptosis. *Mol Endocrinol* 2003; 17: 161–171.
 74. Rosenstock J, Foley JE, Rendell M, et al. Effects of the dipeptidyl peptidase-IV inhibitor vildagliptin on incretin hormones, islet function, and postprandial glycemia in subjects with impaired glucose tolerance. *Diabetes Care* 2008; 31: 30–35.
 75. Young A. Tissue expression and secretion of amylin. *Adv Pharmacol* 2005; 52: 19–45.
 76. Baron AD, Kim D, Weyer C. Novel peptides under development for the treatment of type 1 and type 2 diabetes mellitus. *Curr Drug Targets Immune Endocr Metabol Disord* 2002; 2: 63–82.
 77. Ratner RE, Want LL, Fineman MS, et al. Adjunctive therapy with the amylin analogue pramlintide leads to a combined improvement in glycemic and weight control in insulin-treated subjects with type 2 diabetes. *Diabetes Technol Ther* 2002; 4: 51–61.
 78. Smith SR, Aronne LJ, Burns CM, et al. Sustained weight loss following 12-month pramlintide treatment as an adjunct to lifestyle intervention in obesity. *Diabetes Care* 2008; 31: 1816–1823.
 79. Drucker DJ. The biology of incretin hormones. *Cell Metab* 2006; 3: 153–165.
 80. Wookey PJ, Lutz TA, Andrikopoulos S. Amylin in the periphery II: an updated mini-review. *Sci World J* 2006; 6: 1642–1655.

Nutritional Strategies in Type 2 Diabetes

Nichola Davis, MD, MS, Bernice Forbes, MD, and Judith Wylie-Rosett, EdD, RD

Albert Einstein College of Medicine, Bronx, NY

ABSTRACT

Weight loss is a key goal in the management of patients with type 2 diabetes mellitus. There are conflicting weight-loss strategies, including low-carbohydrate diets, low glycemic-index diets, low-fat vegan diets, conventional low-fat diets, and high-protein and high-monounsaturated fat diets. There is no evidence to suggest that 1 nutritional strategy is clearly the best. The most successful strategy is one that the patient can adopt and follow in the long term. *Mt Sinai J Med* 76:257–268, 2009. © 2009 Mount Sinai School of Medicine

Key Words: conventional low-fat diets, high-monounsaturated fat diets, high-protein diets, low-carbohydrate diets, low-fat vegan diets, low-glycemic index diets, nutritional strategy, type 2 diabetes mellitus, weight loss.

The prevalence of diabetes in the United States is predicted to increase from the estimate of 20 million in 2005 to 48.3 million by 2050.^{1,2} Although prevalence reporting does not distinguish types of diabetes, type 2 diabetes mellitus (T2DM) accounts for the vast majority of cases. The projected doubling of diabetes is closely linked to the dramatic projected increase in the prevalence of overweight and obesity. The prevalence of overweight and obesity in the United States is predicted to increase from 65% to 80% during the next 15 years and reach 100% by the year 2048.³

The insulin resistance associated with excess body weight complicates the management of diabetes for the majority of patients with T2DM. Defects in insulin action, secretion, or both lead to chronic hyperglycemia and disturbances in the

metabolism of carbohydrates, proteins, and fat.⁴ Hyperglycemia both causes acute symptoms of diabetes and results in microvascular complications such as neuropathy, retinopathy, and nephropathy. In particular, postprandial hyperglycemia increases the severity of diabetes and leads to macrovascular complications such as stroke and cardiovascular disease (CVD), which is the leading cause of death in diabetes.⁵

In this article, we review the role of diet in the development and management of diabetes. We focus on the benefits of modest weight reduction in diabetes treatment and on strategies to achieve weight reduction. Additionally, we examine the effects of dietary composition on weight reduction and glycemic control.

DIET AND DEVELOPMENT OF DIABETES

There have been numerous studies that examine the relationship between the type of diet and the incidence of T2DM. Excess caloric intake and obesity predict the incidence of diabetes.⁶ The Nurses' Health Study followed 84,941 female nurses for 16 years and found that although the risk of diabetes was significantly increased by a poor diet, smoking, lack of exercise, and abstinence from alcohol, overweight and obesity were the most important predictors of diabetes.⁷ Prevention studies have demonstrated that lifestyle interventions including modest weight reduction and physical activity result in a dramatic reduction in the incidence of diabetes.⁸ The Diabetes Prevention Program demonstrated a 58% reduction in the incidence of diabetes among patients at risk for diabetes who were randomized to an intensive lifestyle intervention (ILI) for 3.5 years. Table 1 highlights key features of the intervention components of the Diabetes Prevention Program. A reduction in the risk of T2DM is also associated with diets high in whole grains, cereal fiber, and dietary magnesium.^{9,10} Diet is clearly an important factor in preventing diabetes, but what is known about the role of diet in treating T2DM?

Address Correspondence to:

Nichola Davis

Albert Einstein College of
Medicine
Bronx, NY

Email: ndavis@montefiore.org

Table 1. *Overview of Interventions Used in the Diabetes Prevention Program and Look AHEAD.*

Intervention Components	Diabetes Prevention Program	Look AHEAD Trial
Individual goals	Lose 7% of initial weight: 1200 cal/day for 120–170 lb, 1500 cal/day for 175–215 lb, 1800 cal/day for 220–245 lb, and 2000 cal/day for >250 lb	Lose 10% of initial weight: 1200–1500 cal/day for <250 lb and 1500–1800 cal/day for ≥ 250 lb; 175 minutes of physical activity per week
Core curriculum (first 6 months)	16 sessions (individual sessions)	24 sessions (mixture of group and individual sessions)
Follow-up intervention after core curriculum	Consultation (individual) at least every 2 months	18 sessions (mixture of group and individual sessions) for months 7–12; monthly individual sessions for years 2–4
Self-monitoring	Counting fat grams (25% of calories/day goal); calorie counting added if needed to reach weight goal; monitoring of weekly minutes of physical activity	Calorie counting with inclusion of fat gram counting; portion control; monitoring of weekly minutes of physical activity
Assessment of barriers and toolbox	Assessment of barriers; problem solving as a basic toolbox approach at onset; use of budget to address barriers after completion of core	Assessment of barriers; problem solving as a basic toolbox approach at onset; motivational interviewing; behavioral contracts; use of budget after completion of core to achieve or maintain goal weight
Campaigns or refresher classes (3–4 sessions)	Offered 2–3 times a year after completion of core; theme-focused; contest included to enhance motivation	Same as the Diabetes Prevention Program starting at year 2
Meal replacements	Toolbox and campaign strategy after core	Used as a strategy to achieve portion control
Study-sponsored physical activity sessions	Two times per week throughout the study	Periodically in campaigns

NOTE: This table was adapted from the *Journal of the American Dietetic Association*.¹⁸

Abbreviation: AHEAD, Action for Health in Diabetes.

TREATMENT OF DIABETES

Glycemic control is a key outcome in diabetes management. The United Kingdom Prospective Diabetes Study demonstrated that a 1% decrease in hemoglobin A1c (HbA1c) in patients with T2DM resulted in a 21% reduction in any diabetes-related endpoint: 21% for diabetes-related death, 14% for myocardial infarction, and 37% for microvascular complications.¹¹ Although the United Kingdom Prospective Diabetes Study achieved glycemic control through medications, diet and weight reduction can also improve glycemic control.

WEIGHT LOSS AND MAINTENANCE

Several studies to date have examined the effects of behavioral interventions and caloric restriction in achieving weight loss and glycemic control.^{12,13} Most studies are short-term and last less than 1 year. In general, the degree of weight loss correlates with the degree of glycemic control.¹⁴ Wing¹² demonstrated that the average weight reduction with behavioral

interventions was 9 kg at 4 to 6 months, and this was associated with a reduction of 1.67% in HbA1c. In many behavioral interventions, there is weight regain after the initial 6-month weight-loss period. Wing further demonstrated that at 1 year, the average weight loss was 6.4 kg, and this was associated with a 0.55% lowering of HbA1c. It is important to recognize that the reductions obtained through weight loss are comparable to those achieved with oral hypoglycemic monotherapy, which on average reduces HbA1c by 1%.¹⁵

If patients stop taking their oral hypoglycemics, glycemic control in general will worsen. Similarly, when patients achieve HbA1c reduction through weight loss, glycemic control can worsen with weight regain. Therefore, it is important to recognize and understand the barriers to maintaining weight loss. The challenge of successful weight loss is highlighted in a study from 1959: of 100 obese individuals who were followed up 2 years after referral to a nutritional weight-loss program, only 2% of the subjects maintained a weight loss of 20 lb or more.¹⁶ According to Wing and Hill,¹⁶ the definition of successful weight-loss maintainers is “individuals

who have intentionally lost at least 10% of their body weight and kept it off at least 1 year.” The National Weight Control Registry is a cohort of approximately 4000 adults who have lost at least 30 pounds and kept it off for a year.¹⁷ Data from this registry highlight key behaviors that successful weight-loss maintainers use. These 6 strategies include (1) a low-calorie, low-fat diet, (2) eating breakfast, (3) consistent patterns of eating, (4) high levels of physical activity, (5) regular self-monitoring of weight, and (6) preventing weight regains from turning into big relapses.¹⁶ There is, however, a selection bias in participants enrolled in the National Weight Control registry, in that participants are self-selected and are 95% Caucasian, 82% college-educated, and 77% female.¹⁷ Successful strategies for weight-loss maintenance in more diverse populations have not been well studied.

Although the short-term benefits of weight loss for glycemic control have been well studied, the effects of long-term weight loss on T2DM are currently being examined in the Look AHEAD (Action for Health in Diabetes) study.¹⁴ Look AHEAD is a long-term, multicenter trial of 5145 patients with T2DM randomly assigned to either an ILI or enhanced usual care consisting of diabetes support and education. It started in 2001 and is scheduled to end in 2012. Its main objective is to evaluate whether long-term weight loss achieved through an ILI that includes diet, exercise, and behavior modification can reduce cardiovascular morbidity and mortality in T2DM.¹⁴ The main goals of the intervention are to achieve a mean initial weight loss $\geq 7\%$ and to increase moderately intense physical activity to ≥ 175 minutes/week.¹⁴ To achieve this weight loss, Look AHEAD used an intensive dietary intervention that included a reduced-calorie, low-fat diet and meal replacements and allowed selected weight-loss medications. Table 1 further highlights the components of the Look AHEAD intervention.

Results of this trial at 1 year for the ILI group showed weight loss of 8.6% versus 0.7% in the diabetes support and education group and a reduction in HbA1c from 7.3% to 6.6% in the ILI arm versus 7.3% to 7.2% in the control arm.¹⁸ Additionally, the ILI arm demonstrated a greater decrease in diabetes medications, antihypertensives, and cholesterol-lowering medications. The goal of $<7\%$ HbA1c was met by 72.7% in the ILI arm versus 23.6% in the control arm.¹⁸ The large weight loss achieved in Look AHEAD may be due to the group and individual contact, more intensive physical activity goals, and strict dietary intervention. Earlier studies by Williams *et al.*¹⁹ and Wing¹² found similar results in T2DM subjects at 1 year, showing better glycemic control and improved cardiovascular risk

factors with weight loss. Definitive effects of the maintenance and long-term effects of this weight loss on improving cardiovascular risk factors and decreasing cardiovascular events are not well known.

EFFECTS OF NUTRIENT COMPOSITION ON WEIGHT AND GLYCEMIC CONTROL

The macronutrients—carbohydrate, protein, and fat—have varying effects on blood glucose levels. Dietary carbohydrates, which elevate blood glucose, have the greatest influence on blood glucose. Fats minimally affect insulin secretion and do not significantly affect blood glucose. Protein improves insulin secretion and may lower blood glucose levels.²⁰ Altering the intake of any of these macronutrients will result in a concomitant change in other macronutrients if energy intake is unchanged. For example, a low-carbohydrate diet will have a higher proportion of calories from fat, protein, or both.

HISTORY OF DIETARY RECOMMENDATIONS IN DIABETES

For the past 3500 years, the optimal nutrient composition for treating diabetes has been widely debated. As early as 1550 BC, the first diabetes dietary treatments were listed in the Ebers Papyrus, and they focused on the consumption of foods rich in carbohydrates. In the 6th century AD, the notion of carbohydrate restriction emerged. Recommendations during the 17th and 18th centuries ranged from the consumption of meat and fat and avoidance of carbohydrates to the consumption of a high-carbohydrate diet to replace sugar loss. Measured diabetic diets limiting carbohydrates and fasting became more popular during the 19th and early 20th centuries, and these were the mainstay of treatment at treatment centers before the discovery of insulin.²¹

Since the 1950 s, the American Diabetes Association (ADA) has consulted a panel of experts to formulate recommendations for diabetes nutrition management. The initial recommendations were to reach an ideal body weight, avoid simple sugars, and individualize programs with an exchange system. This exchange system was designed to be used for meal planning to achieve a macronutrient distribution of 40% carbohydrates, 40% fat, and 20% protein. As knowledge of diabetes increased and more complications arose such as CVD, the focus of recommendations shifted to reduce dietary fat

Table 2. Sample Meal Plans with Different Nutritional Strategies.

Low-Carbohydrate Meal	Low-Fat Meal	Low-Glycemic Index Meal
5 oz of baked chicken breast in cream sauce 1 cup of broccoli with butter 1 cup of lettuce and 1 cup of sliced cucumber with Caesar salad dressing (2 tablespoons) 8 oz of a diet drink Macronutrient composition: 480 cal 65% fat (34 g of fat) 19% protein 16% carbohydrate	4 oz of grilled chicken breast with teriyaki sauce 1 cup of broccoli no butter 1/2 cup of lettuce and 1/2 cup of sliced cucumber with vinaigrette dressing (2 tablespoons) 1/2 cup of white rice 8 oz of a diet drink Macronutrient composition: 480 cal 23% fat (12 g of fat) 37% protein 40% carbohydrate	4 oz of grilled chicken breast 1/2 cup of steamed broccoli with no butter 1/2 cup of lettuce and 1/2 cup of sliced cucumber with vinaigrette dressing (2 tablespoons) 1/2 cup of parboiled rice 1/2 cup of beans in a tomato sauce 8 oz of a diet drink Macronutrient composition: 485 cal 23% (12 g fat) 36% protein 41% carbohydrate

and increase carbohydrates in efforts to reduce CVD complications. In 1994, the focus centered on individualized management of patients based on a dietitian's assessment. Currently, there are no specific recommendations for the balance between the ingestion of total fat and carbohydrate intake.²¹

The conventional dietary strategy, a low-fat diet, has been contested as there is mounting evidence regarding potential effects of other dietary strategies on glycemic control. Because of the many controversies and debates surrounding the dietary management of T2DM, the goal of this article is to summarize the current literature on different nutritional strategies and present a comprehensive review that can assist healthcare providers in the management of their patients. Included in this review are observational and randomized controlled clinical studies of adults with T2DM in which diet was an intervention and glycemic control was a primary outcome. The dietary strategies that we review are low-carbohydrate diets, low-glycemic index (GI) diet, fat restriction, a vegetarian diet, a high-monounsaturated fat (MUFA) diet, and high-protein diet. Table 2 provides an example of low-carbohydrate, low-fat, and low-GI meals, all with similar caloric contents and with some similarities in foods.

MODIFICATION OF CARBOHYDRATE INTAKE IN DIETARY STUDIES

Compared to the other macronutrients, carbohydrates provide the majority of glucose for metabolism. Both the quantity and source of carbohydrates in food influence postprandial glucose levels. Different factors intrinsic to food can exert an influence on blood glucose, such as the physical form of the food, ripeness, degree of processing, type of starch, and preparation methods. Extrinsic variables such as the coingestion of protein and fat, previous

food intake, fasting blood glucose, and degree of insulin resistance will also influence the effect of a specific carbohydrate-containing food on blood glucose levels.⁵ Because carbohydrates have a major influence on glucose levels, low-carbohydrate diets have been espoused in an attempt to reduce postprandial glucose. In addition to lowering blood glucose, carbohydrate restriction leads to a significant reduction in insulin resistance, a reduction in levels of insulin, and utilization of stored fat to fuel the body's energy needs.⁵

As shown in Table 3, several studies have demonstrated that there may be beneficial effects of a low-carbohydrate diet on glycemic control.^{22–27} McAuley *et al.*²⁸ compared the effects of the high-fat Atkins diet, the high-protein Zone diet, and the high-carbohydrate, high-fiber diet in 96 insulin-resistant women for 8 weeks. The results showed a significantly greater decrease in body weight, waist circumference, and triglycerides in the high-fat Atkins diet and high-protein Zone diet, suggesting that these popular diets may have a more positive impact on insulin resistance than the high-carbohydrate diet.²⁸ Boden *et al.*²² conducted a study of 10 obese patients with T2DM who were observed during 7 days of their usual diet followed by 14 days of a low-carbohydrate intervention in which participants consumed 21 g of carbohydrates daily. After 14 days of the low-carbohydrate diet, participants had improvements in glycemic control and insulin sensitivity, decreases in mean body weight and mean body fat, and a reduction in triglycerides and total cholesterol (Table 3). Long-term effects of a low-carbohydrate diet could not be determined from this study; however, Nielsen and Joensson²³ did a retrospective follow-up of 16 obese T2DM subjects who followed a low-carbohydrate diet (20% carbohydrates) or a control diet (55%–60% carbohydrates) for 22 months. After 22 months, in the low-carbohydrate group, there was greater improvement in HbA1c, and diabetes medications were eliminated or decreased.

This study was not randomized, and participants self-selected either a low-carbohydrate diet or a control diet. However, it demonstrates that in motivated patients who prefer a low-carbohydrate diet, significant and sustained improvements in glycemic control may be achievable.²³

Yancy *et al.*²⁴ studied 21 overweight T2DM subjects for 16 weeks on a low-carbohydrate diet (≤ 20 g/day). Following the low-carbohydrate diet, the subjects experienced decreases in HbA1c, fasting plasma glucose, body weight, and fasting serum triglycerides. Antihyperglycemic medications were reduced in 10 subjects, discontinued in 7, and unchanged in 4. Although this study was limited by its lack of a control group, its small sample size, and its short duration, the significant improvement in HbA1c in the setting of reduced diabetes medications is impressive.

There are no long-term randomized studies that have specifically examined glycemic control in patients with T2DM on low-carbohydrate and low-fat diets, and therefore recommendations about this dietary strategy cannot be made. The primary outcomes for most of the randomized studies comparing low-carbohydrate diets to low-fat diets have been weight and lipids. In 1 study, 132 obese adults were randomized to a low-carbohydrate or low-fat diet; 39% ($n = 52$) had T2DM.²⁵ After 6 months, the patients with T2DM in the low-carbohydrate diet group had a greater decrease in HbA1c. Those subjects who lost more than 5% of their initial weight also experienced a greater improvement in insulin sensitivity, greater weight reduction, and improved triglyceride levels.²⁵ Daly *et al.*²⁶ specifically examined glycemic control in 102 obese patients with T2DM randomized to a low-carbohydrate or low-fat diet. After 3 months, both dietary groups experienced significant decreases in HbA1c without a significant difference between the dietary groups. Participants in the low-carbohydrate group also had a greater reduction in body weight and an improved total cholesterol to high-density lipoprotein (HDL) ratio.²⁶

More recently, Shai *et al.*²⁷ randomized 322 moderately obese subjects, 36 of whom had diabetes, to a calorie-restricted low-fat, a non-calorie-restricted, low-carbohydrate diet, or a calorie-restricted Mediterranean diet for 2 years. After 2 years, there was a significant decrease in HbA1c of $0.9\% \pm 0.8\%$ in the low-carbohydrate group, a significant decrease in fasting plasma glucose of 32.8 mg/dL in the Mediterranean diet group versus an increase in the low-fat group, decreased insulin levels in all groups, and a greater decrease in insulin resistance in the Mediterranean diet group.²⁷ Although this study was longer

than most dietary studies, the sample of patients with diabetes was very small.

Collectively, the data from low-carbohydrate diets suggest beneficial effects on glycemic control. Most studies, however, have small sample sizes and are short-term in duration; this limits the recommendations that can be made.

There are some concerns about low-carbohydrate diets in diabetes, including the importance of nutrients, water-soluble vitamins, minerals, and fiber found in some carbohydrate-containing foods.⁵ However, adverse effects of low-carbohydrate diets are largely unfounded. On the basis of the authors' experiences, patients with T2DM who prefer to try a low-carbohydrate dietary strategy should be closely monitored by their physician for the need for medication changes. In particular, medication reduction is necessary to prevent hypoglycemia when low-carbohydrate diets are initiated.

GLYCEMIC INDEX

GI is a measure of the changes in blood glucose after the consumption of carbohydrate-containing foods. It ranks foods according to their effect on postprandial glycemia. It quantifies the blood glucose response to carbohydrates present in different foods and compares them on a weight-for-weight basis. GI is the increase in blood glucose (over the fasting level) observed 2 hours after the consumption of a set amount of carbohydrates from an individual food item.^{29,30} This value is compared to the response to a reference food such as glucose or white bread containing an equivalent amount of carbohydrates. To improve the reliability of evaluating the glycemic response of a diet, the use of glycemic load has been suggested. Glycemic load is the product of GI and the amount of carbohydrates in a serving. Knowing the glycemic load of individual foods in a meal allows the calculation of the overall glycemic load of a meal.⁵ Foods with a low GI lower postprandial hyperglycemia³¹; low-GI diets contain mostly carbohydrates from low-GI sources such as parboiled rice, barley, oats, pumpnickel, beans, peas, lentils, pasta, and bulgur. High-GI diets contain white bread, potatoes, wheat meal, sweetened breakfast cereals, and rice.¹¹

Table 3 highlights the studies that have examined modifications in the dietary GI as an intervention. Brand-Miller *et al.*¹¹ conducted a meta-analysis of randomized controlled trials conducted between 1981 and 2001 that evaluated the effect of GI on blood glucose control. Fourteen studies involving 356 subjects were included. Primary outcomes were HbA1c,

Table 3. *Nutritional Strategies with CHO Modification.*

Study	Intervention	n	Duration	Outcome Measures: Effects (Results)
Boden <i>et al.</i> ²²	Observational Low-CHO diet (21 g/day) versus the usual diet	10 obese T2DM subjects	7 days for the usual diet and 14 days for the low-CHO diet	On the low-CHO diet: HbA1c: 7.3%–6.8% Weight loss: 2.02 kg versus no change in the usual diet Insulin sensitivity: 75% improvement Triglycerides: reduced by 35% Total cholesterol: reduced by 10%
Nielsen and Joensson ²³	Retrospective follow-up Low-CHO diet (20% CHO, 30% protein, and 50% fat) versus a control (55%–60% CHO, 15% protein, and 25%–30% fat)	16 obese T2DM subjects	22 months	On the low-CHO diet: HbA1c: 8.0%–6.9% Weight loss: 8.6 kg
Yancy <i>et al.</i> ²⁴	Observational Low-CHO diet (≤20 g/day)	21 overweight T2DM subjects	16 weeks	Significant reductions: HbA1c: 7.5%–6.3% Weight: 8.7 kg Triglycerides: reduced by 42%
Samaha <i>et al.</i> ²⁵	RCT Low-CHO diet (≤30 g/day) versus a low-fat diet (≤30% fat)	132 obese subjects (39% with T2DM)	6 months	Significant reductions in the low-CHO group versus the low-fat group: HbA1c: 7.8%–7.2% in the low-CHO group versus no change in the low-fat group Weight: 5.8 kg in the low-CHO group versus 1.9 kg in the low-fat group Triglycerides: 20% reduction in the low-CHO group versus a 4% reduction in the low-fat group
Daly <i>et al.</i> ²⁶	RCT Low-CHO diet (up to 70 g/day) versus a low-fat diet	102 obese T2DM subjects	3 months	Greater reductions in the low-CHO group: HbA1c: absolute change of 0.55 in the low-CHO group versus 0.23 in the low-fat group Weight: 3.55 kg in the low-CHO group versus 0.92 kg in the low-fat group Total cholesterol-to-HDL ratio: 0.48 in the low-CHO group versus 0.10 in the low-fat group
Shai <i>et al.</i> ²⁷	RCT Low-CHO diet (20 g/day) versus the Mediterranean diet [maximum of 35% fat; main sources: 30–45 g of olive oil and 5–7 nuts (<20 g/day)] versus a low-fat diet (30% fat, 10% saturated fat, and 300 mg of cholesterol/day)	322 obese subjects (36 with T2DM)	24 months	In subjects with diabetes: HbA1c: significant reduction of 0.9% in the low-CHO group Weight loss: 4.7 kg in the low-CHO group versus 4.4 kg in the Mediterranean group versus 2.9 kg in the low-fat group (all groups lost weight, but greater reductions occurred in the Mediterranean and low-CHO groups versus the low-fat group)
Brand-Miller <i>et al.</i> ¹¹	Retrospective meta-analysis of randomized crossover or parallel experimental design Low-GI diet versus a high-GI diet	14 studies, 356 subjects	12 days to 12 months	Significant reductions in the low-GI group versus the high-GI group: HbA1c: 0.43% in the low-GI group versus the high-GI group Fructosamine: 0.18 mmol/L in the low-GI group versus the high-GI group

(continued overleaf)

Table 3. (Continued).

Study	Intervention	n	Duration	Outcome Measures: Effects (Results)
Rizkalla <i>et al.</i> ³²	Randomized crossover trial Low-GI diet (GI <45) versus a high-GI diet (GI > 60)	12 T2DM males	12 weeks	Greater reductions in the low-GI group versus the high-GI group: HbA1c: 7.56%–7.17% in the low-GI group versus 7.45%–7.57% in the high-GI group Weight loss: no effect Total cholesterol: 0.87 mmol/L in the low-GI group versus 0.11 mmol/L in the high-GI group LDL: 0.83 mmol/L in the low-GI group versus 0.14 mmol/L in the high-GI group
Heilbronn <i>et al.</i> ³³	RCT High-GI diet versus a low-GI diet (both consisted of 60% CHO, 20% protein, and 15% fat)	45 subjects (23 males and 22 females) diagnosed with T2DM in the past 10 years	12 weeks with 4 weeks before the study	HbA1c: 6.65%–6.04% in the low-GI group versus 6.35%–6.06% in the high-GI group Weight loss: 4.4 kg in the low-GI group versus 4.8 kg in the high-GI group LDL cholesterol: 18% reduction in the low-GI group with low glucose tolerance versus 5% in the high-GI group
Jenkins <i>et al.</i> ³⁵	Randomized parallel study Low-GI diet versus a high-cereal fiber diet	210 T2DM subjects taking antihyperglycemic medications	6 months	HbA1c: greater reduction in the low-GI group (7.14%–6.64%) versus the high-cereal fiber group (7.07%–6.89%) Weight loss: no significant reduction HDL cholesterol: increased (from 41.9 to 43.6 mg/dL) in the low-GI group and decreased (from 43.1 to 42.8 mg/dL) in the high-cereal fiber group
Ma <i>et al.</i> ³⁴	RCT Low-GI diet versus the ADA diet	40 poorly controlled T2DM subjects	12 months	HbA1c: 8.74%–8.39% in the low-GI group versus 8.10%–7.67% in the ADA group Total cholesterol: 1.95 mg/dL in the low-GI group versus 18.39 mg/dL in the ADA group (lower likelihood of increasing medication dosage or switching to another drug; odds ratio, 0.26)

Abbreviations: ADA, American Diabetes Association; CHO, carbohydrate; GI, glycemic index; HbA1c, hemoglobin A1c; HDL, high-density lipoprotein; LDL, low-density lipoprotein; RCT, randomized clinical trial; T2DM, type 2 diabetes mellitus.

which reflects glycemic control over the past 6 to 12 weeks, and fructosamine, which reflects glycemic control over the past 2 to 4 weeks. Results showed that after an average duration of 10 weeks, there was a greater reduction in HbA1c and fructosamine in the low-GI diet compared to the high-GI diet.¹¹ The meta-analysis was limited by the short duration of the studies, the small number of subjects, and uncertainty regarding dietary compliance in free-living populations, and the authors concluded that the consumption of low-GI foods had a small beneficial effect on glycemic control.

Subsequent to this meta-analysis, there have been several additional studies comparing low-GI and high-GI diets. A short-term crossover study was conducted in 12 men with T2DM randomly allocated to either a high-GI diet or a low-GI diet for 4 weeks each and then crossed over to the other group with a 2-week washout period in between.³² Compared to the high-GI diet, the low-GI diet resulted in a greater

reduction in HbA1c and fasting plasma glucose and an improvement in insulin sensitivity. During the low-GI period, there was a significant decrease in total cholesterol and low-density lipoprotein (LDL; Table 1). Although the extrapolation of study findings is limited by the homogeneous population of men studied, it demonstrates that low-GI diets may lead to significant improvements in glycemic control and lipid profiles.³² This knowledge may help guide healthy subjects to begin to choose low-GI foods and decrease their risk for metabolic diseases and CVDs. Heilbronn *et al.*³³ randomized 45 overweight subjects diagnosed with T2DM in the past 10 years to a high-GI diet or a low-GI diet and found a greater but not statistically significant decrease in HbA1c with the low-GI diet. Ma *et al.*³⁴ randomized 40 poorly controlled T2DM subjects to a low-GI diet or what is considered the traditional ADA dietary approach, which emphasizes carbohydrate counting and 55% of total energy from carbohydrates for 12 months.

The intervention consisted of 8 educational sessions. Although both groups had significant improvements in HbA1c, there were no significant differences between groups. Members of the low-GI group, however, had lower LDL cholesterol and were less likely to increase their diabetes medications. This suggests that a low-GI diet may be a viable alternative to the traditional ADA diet in controlling hyperglycemia and may allow for a reduction in diabetes medication use.

Recently, Jenkins *et al.*³⁵ randomized 210 subjects taking antihyperglycemic medications to a low-GI diet or a high-cereal fiber diet for 6 months. There was a greater reduction in HbA1c and an increase in HDL cholesterol in the low-GI group. The high-cereal fiber group had less reduction in HbA1c and a decrease in HDL. There was no change in weight in either dietary group (Table 3). The results of this study suggest that low-GI diets may improve both glycemic control and some cardiovascular risk factors.

Collectively, these studies suggest additional benefits with the use of GI and glycemic load in controlling postprandial glycemia in comparison with evaluating only total carbohydrates. Low-GI foods are generally rich in fiber and other nutrients, and increased consumption of these foods may be beneficial for patients with T2DM.

MODIFICATION OF FAT INTAKE

Excess body fat is the single most important determinant of T2DM. Dietary saturated fat intake has been associated with the development of T2DM.⁵ Additionally, patients with T2DM are at high risk for CVD, and this has resulted in dietary recommendations for a low-fat diet to reduce CVD risk.³¹ Interventions aimed at reducing fat intake, however, are not always accompanied by reductions in weight and HbA1c. Table 4 highlights interventions that have modified fat intake.

Low-fat, plant-based diets have been advocated for their effective control of postprandial glycemia and a reduction in medications and neuropathic symptoms in T2DM patients.³⁰ Vegan diets also improve lipid profiles because of the low amounts of saturated fat and dietary cholesterol and the presence of fiber and plant constituents that help to lower cholesterol.³⁷ Nicholson *et al.*³⁶ studied 11 T2DM subjects randomly assigned to either a low-fat vegan diet (7 subjects) or a conventional low-fat diet (4 subjects) for 12 weeks. Data from the study revealed no significant effect on HbA1c with either diet or on total cholesterol, LDL, or triacylglycerol. There was

a greater reduction in body fasting plasma glucose and HDL with the low-fat vegan diet (see Table 4.) Barnard *et al.*³⁷ randomized 99 T2DM subjects to a low-fat vegan diet ($n = 49$) or the ADA diet ($n = 50$) for 22 weeks. Participants in the low-fat vegan diet had more significant reductions in HbA1c, cholesterol, urinary albumin, and body weight. The change in body weight was found to be significantly associated with changes in HbA1c, and each kilogram of weight loss was associated with a decrease in HbA1c of 0.12%.³⁷ These studies demonstrate that a low-fat vegan diet and an ADA diet are both effective in glycemic and lipid control in T2DM subjects, but greater benefits may be seen with a low-fat vegan diet.

In addition to the more conventional low-fat diet, higher fat diets that focus on the type of fat consumed have also been studied. A diet high in MUFA has been proposed by some because it increases adipose tissue metabolism and may facilitate weight loss. The specific weight-loss mechanism of high-MUFA diets is its higher oxidation rate in comparison with saturated fat.³⁸ Diets high in MUFA have demonstrated improvements in metabolic control in some studies and may have differing effects on weight. A meta-analysis of short-term randomized crossover trials compared the effects of a high-MUFA diet and a high-carbohydrate diet in T2DM subjects.³⁹ Nine studies involving 133 patients were included. Dietary interventions ranged in length from 2 to 6 weeks. Results of this meta-analysis revealed no significant effects of either diet on HbA1c or fructosamine. The MUFA-rich diets demonstrated a greater reduction in fasting glucose, and some studies have additionally demonstrated significant reductions in fasting plasma triacylglycerol, very low density lipoprotein cholesterol, and plasma total cholesterol. Results of this meta-analysis suggest that a diet high in MUFA may improve fasting glucose and lipoprotein profiles of T2DM subjects. In contrast to these findings are those of Gerhard *et al.*,⁴⁰ who did not find any difference in HbA1c or cholesterol in patients with diabetes randomized to a low-fat, high-carbohydrate diet or a high-MUFA diet for 6 weeks. Gerhard *et al.* also demonstrated greater weight loss in patients on the low-fat diet versus patients on the high-MUFA diet (Table 4).

Another type of fatty acid that may have a role in the dietary management of T2DM is oleic acid. In studies evaluating its effects, it was found to reduce LDL without decreasing HDL and to provide protection from LDL oxidation, which is associated with atherogenesis.⁴¹ Higashi *et al.*⁴¹ assigned 8 subjects to either a low-fat oleic acid-enriched diet or a low-fat linoleic acid-enriched diet for 3 weeks each in a randomized crossover design.

Table 4. *Nutritional Strategies with Fat Modification.*

Study	Intervention	n	Duration	Outcome Measures: Effects (Results)
Nicholson <i>et al.</i> ³⁶	RCT Low-fat vegan diet (10%–15% protein, <10% fat, and the rest from unrefined complex CHO) versus a conventional low-fat diet (55%–60% CHO, <30% fat, and 200 mg of cholesterol/day)	11 T2DM subjects (7 in the low-fat vegan group and 4 in the low-fat group)	12 weeks	In both diets, no significant effects. HbA1c, total cholesterol, LDL, or triacylglycerol Weight loss: 7.2 kg in the low-fat vegan group versus 3.8 kg in the conventional low-fat group. HDL: 0.2 versus 0.02 mmol/L
Barnard <i>et al.</i> ³⁷	RCT Low-fat vegan diet (10% fat, 15% protein, and 75% CHO) versus the ADA diet (<7% saturated fat, 15%–20% protein, 60%–70% CHO and MUFA, and ≤ 200 mg of cholesterol/day)	99 patients (49 in the low-fat vegan group and 50 in the ADA group)	22 weeks	HbA1c: 8%–7.1% in the low-fat vegan group versus 7.9%–7.4% in the ADA group HbA1c in subjects whose medications were unchanged: 8.07%–6.84% in the low-fat vegan group versus 7.88%–7.5% in the ADA group Weight loss: 5.8 kg in the low-fat vegan group versus 4.3 kg in the ADA group
Garg ³⁹	Meta-analysis of randomized crossover trials High-CHO diet (15%–21% protein, 49%–60% CHO, 20%–32% fat, and 7%–13% MUFA) versus a high-MUFA diet (14%–21% protein, 35%–40% CHO, 37%–50% fat, and 22%–33% MUFA)	9 studies, 133 T2DM subjects	2–6 weeks	No significant effects on HbA1c, fructosamine, LDL On the high-MUFA diet, significant reductions: Fasting plasma triacylglycerol: 19% reduction Total cholesterol: 3% reduction HDL: 41% increase in the high-MUFA group
Gerhard ⁴⁰	RCT Low-fat diet (20% fat, 65% CHO, and 10% refined sugar) versus a high-MUFA diet (40% fat, 45% CHO, and 10% refined sugar)	11 T2DM patients	12 weeks with a washout period of 6–12 weeks between dietary periods	Weight loss: 1.53 kg in the low-fat group versus 0.47 kg in the high-MUFA group No differences between the 2 groups in HbA1c or fructosamine In both groups, reductions in values that failed to reach statistical significance: Total cholesterol: 9.6 mg/dL in the low-fat group versus 10.2 mg/dL in the high-MUFA group LDL cholesterol: 10.2 mg/dL in the low-fat group versus 17.5 mg/dL in the high-MUFA group HDL: 7.1 mg/dL in the low-fat group versus 4.5 mg/dL in the high-MUFA group
Higashi <i>et al.</i> ⁴¹	Randomized crossover 25% fat, 58% CHO, and 17% protein in both diets: an oleic acid-enriched low-fat diet (5% saturated fat, 15% MUFA, and 5% PUFA) versus a linoleic acid-enriched low-fat diet (5% saturated fat, 5% MUFA, and 15% PUFA)	8 T2DM subjects (2 males and 6 females)	3 weeks for each phase	Significant reductions on the linoleic acid-enriched diet versus the oleic acid-enriched diet: Total cholesterol: 4.68 versus 4.69 mmol/L LDL cholesterol: 2.96 versus 2.89 mmol/L

Abbreviations: ADA, American Diabetes Association; CHO, carbohydrate; HbA1c, hemoglobin A1c; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MUFA, monounsaturated fat; PUFA, polyunsaturated fat; RCT, randomized clinical trial; T2DM, type 2 diabetes mellitus.

Table 5. *Nutritional Strategies with Protein Modification.*

Study	Intervention	n	Duration	Outcome Measures: Effects (Results)
Gannon <i>et al.</i> ⁴⁵	Randomized crossover High-protein diet (30% protein, 40% CHO, and 30% fat) versus a control diet (15% protein, 55% CHO, and 30% fat)	12 T2DM patients (10 males and 2 females)	10 weeks with a washout period of 2–5 weeks between the 2 dietary periods	HbA1c: 8.1%–7.3% in the high-protein group versus 8.0%–7.7% in the control group Fasting triacylglycerol: 161 mg/dL in the high-protein group versus 199 mg/dL in the control group No significant change in body weight, insulin, or cholesterol levels
Sargrad <i>et al.</i> ⁴⁶	RCT High-protein diet (27%) versus a high-CHO diet (51%)	12 T2DM patients	8 weeks	Significant reductions in the high-CHO group versus the high-protein group: HbA1c: 8.2%–6.9% in the high-CHO group versus 7.6%–6.6% in the high-protein group HDL: 0.05 mmol/L in the high-CHO group versus 0.08 mmol/L in the high-protein group
Pomerleau <i>et al.</i> ⁴⁷	Randomized crossover High-protein diet (2.0 g/kg of desirable weight per day) versus a moderate-protein diet (0.8 g/kg of desirable weight per day)	12 T2DM subjects	3 weeks each with a 3-week washout between the 2 dietary periods	Significant reductions in both groups: Body weight: 1% reduction in the high-protein group versus a 2% reduction in the moderate-protein group

Abbreviations: CHO, carbohydrate; HbA1c, hemoglobin A1c; HDL, high-density lipoprotein; RCT, randomized clinical trial; T2DM, type 2 diabetes mellitus.

All subjects underwent an oral fat-loading test after a 12-hour fast at the end of each dietary period to assess postprandial lipemia. The results demonstrated a significant decrease in fasting blood glucose after both diets and a decrease in total and LDL cholesterol in the linoleic acid–enriched diet, suggesting that linoleic acid–enriched diets and oleic acid–enriched diets may decrease fasting glucose in T2DM subjects. Oleic acid–enriched diets, however, also increase postprandial chylomicron remnants, which are associated with atherosclerosis.⁴¹ As can be seen, many of these studies on fat modification are also short-term with a small sample size, and therefore broader recommendation cannot be made from them. Conventional low-fat diets, however, as part of reduced-calorie diets have been well studied, can result in weight loss, and may help to reduce CVD risk in patients with T2DM.

MODIFICATION OF PROTEIN INTAKE

Diabetes dietary recommendations have mainly been focused on regulating the amounts and types of carbohydrates and fat in the diet. The role of dietary protein with respect to blood glucose levels has been

relegated to the background. Protein ingestion in some studies has been shown to have neutral effects on postprandial glucose,^{42–45} and the role of protein in glycemic control merits further study.

A few studies have examined protein modification, and they are highlighted in Table 5. Compared to a control diet, a high-protein diet given to 12 subjects with mild, untreated T2DM resulted in significant reductions in HbA1c.⁴⁵ In contrast, Sargrad *et al.*⁴⁶ randomized 12 obese T2DM subjects to either a high-protein (27% protein) diet ($n = 6$) or a high-carbohydrate (51% carbohydrates) diet ($n = 6$) for 8 weeks and demonstrated significant decreases in HbA1c and fasting plasma glucose, increased insulin sensitivity, and a small decrease in HDL in the high-carbohydrate group versus the high-protein group. Another study randomly assigned 12 normotensive T2DM subjects with microalbuminuria to a high-protein diet (2.0 g/kg of desirable weight per day) or a moderate-protein diet (0.8 g/kg of desirable weight per day) for a 3-week crossover trial⁴⁷ and demonstrated significant decreases in fasting blood glucose, 60-minute postprandial blood glucose, and body weight with both diets. Further studies are needed in this area to fully understand the role of protein in diabetes.

CONCLUSION

Weight loss is a key strategy in patients with T2DM, and we have reviewed several different nutritional strategies for both weight and glycemic control. Reducing the intake of 1 macronutrient will alter the distribution of macronutrients if energy intake remains constant; for example, reducing either carbohydrate or fat intake will increase the proportion from the other. Low-carbohydrate diets can achieve a reduction in energy intake and weight loss, and restricting carbohydrates may have additional benefits for glycemic control. Low-glycemic diets improve lipid profiles but show conflicting results regarding glycemic control. Low-fat vegan diets also show conflicting results with respect to glycemic control. On the other hand, conventional low-fat diets have a positive effect on weight reduction. Results from high-protein and high-MUFA diets are conflicting. There is no evidence to suggest that 1 nutritional strategy is clearly the better strategy, and the nutritional strategies in diabetes management are not meant to be 1-size-fits-all approaches. The most successful nutritional strategy is one that the patient can adapt and follow over the long term. Therefore, working with patients to follow a strategy that may work best for them and taking into account individual dietary preferences may be the most successful approach of all.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Narayan K, Boyle J, Griss L, et al. Impact of Recent Increase in Incidence on Future Diabetes Burden. *Diabetes Care* 2006; 29: 2114–2116.
- Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. *Phys Ther* 2008; 88: 1254–1264.
- Wang Y, Beydoun MA, Liang L, et al. Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. *Obesity (Silver Spring)* 2008; 16: 2323–2330.
- Nield L, Moore HJ, Hooper L, et al. Dietary advice for treatment of type 2 diabetes mellitus in adults. *Cochrane Database Syst Rev* 2007; CD004097.
- Sheard NF, Clark NG, Brand-Miller JC, et al. Dietary carbohydrate (amount and type) in the prevention and management of diabetes: a statement by the American Diabetes Association. *Diabetes Care* 2004; 27: 2266–2271.
- Wang J, Luben R, Khaw KT, et al. Dietary energy density predicts the risk of clinically incident type 2 diabetes: the EPIC Norfolk study. *Diabetes Care* 2008; 31: 2120–2125.
- Hu FB, Manson JE, Stampfer MJ, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 2001; 345: 790–797.
- Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002; 346: 393–403.
- Fung TT, Hu FB, Pereira MA, et al. Whole-grain intake and the risk of type 2 diabetes: a prospective study in men. *Am J Clin Nutr* 2002; 76: 535–540.
- Meyer KA, Kushi LH, Jacobs DR Jr, et al. Carbohydrates, dietary fiber, and incident type 2 diabetes in older women. *Am J Clin Nutr* 2000; 71: 921–930.
- Brand-Miller J, Hayne S, Petocz P, Colagiuri S. Low-glycemic index diets in the management of diabetes: a meta-analysis of randomized controlled trials. *Diabetes Care* 2003; 26: 2261–2267.
- Wing RR. Behavioral treatment of obesity. Its application to type II diabetes. *Diabetes Care* 1993; 16: 193–199.
- Wing RR, Marquez B. Behavioral aspects of weight loss in type 2 diabetes. *Curr Diab Rep* 2008; 8: 126–131.
- Wadden TA, West DS, Delahanty L, et al. The Look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. *Obesity (Silver Spring)* 2006; 14: 737–752.
- Bolen S, Feldman L, Vassy J, et al. Systematic review: comparative effectiveness and safety of oral medications for type 2 diabetes mellitus. *Ann Intern Med* 2007; 147: 386–399.
- Wing RR, Hill JO. Successful weight loss maintenance. *Annu Rev Nutr* 2001; 21: 323–341.
- Wing RR, Phelan S. Long-term weight loss maintenance. *Am J Clin Nutr* 2005; 82(suppl): 222S–225S.
- Pi-Sunyer X, Blackburn G, Brancati FL, et al. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: one-year results of the Look AHEAD trial. *Diabetes Care* 2007; 30: 1374–1383.
- Williams KV, Mullen ML, Kelley DE, Wing RR. The effect of short periods of caloric restriction on weight loss and glycemic control in type 2 diabetes. *Diabetes Care* 1998; 21: 2–8.
- Gannon MC, Nuttall FQ. Control of blood glucose in type 2 diabetes without weight loss by modification of diet composition. *Nutr Metab (Lond)* 2006; 3: 16.
- Wylie-Rosett J. Diabetes mellitus. In: Bowman BA, ed. *Present Knowledge in Nutrition: Nutrition and Chronic Disease*. 9th ed. Washington, DC: International Life Sciences Institute Press; 2006, chapter 20, 669–686.
- Boden G, Sargrad K, Homko C, et al. Effect of a low-carbohydrate diet on appetite, blood glucose levels, and insulin resistance in obese patients with type 2 diabetes. *Ann Intern Med* 2005; 142: 403–411.
- Nielsen JV, Joensson E. Low-carbohydrate diet in type 2 diabetes. Stable improvement of bodyweight and glycemic control during 22 months follow-up. *Nutr Metab (Lond)* 2006; 3: 22.
- Yancy WS Jr, Foy M, Chalecki AM, et al. A low-carbohydrate, ketogenic diet to treat type 2 diabetes. *Nutr Metab (Lond)* 2005; 2: 34.

25. Samaha FF, Iqbal N, Seshadri P, et al. A low-carbohydrate as compared with a low-fat diet in severe obesity. *N Engl J Med* 2003; 348: 2074–2081.
26. Daly ME, Paisey R, Paisey R, et al. Short-term effects of severe dietary carbohydrate-restriction advice in type 2 diabetes—a randomized controlled trial. *Diabet Med* 2006; 23: 15–20.
27. Shai I, Schwarzfuchs D, Henkin Y, et al. Weight loss with a low-carbohydrate, Mediterranean, or low-fat diet. *N Engl J Med* 2008; 359: 229–241.
28. McAuley KA, Hopkins CM, Smith KJ, et al. Comparison of high-fat and high-protein diets with a high-carbohydrate diet in insulin-resistant obese women. *Diabetologia* 2005; 48: 8–16.
29. Jenkins DJ, Kendall CW, Augustin LS, et al. Glycemic index: overview of implications in health and disease. *Am J Clin Nutr* 2002; 76: 266S–273S.
30. Jenkins DJ, Wolever TM, Taylor RH, et al. Glycemic index of foods: a physiological basis for carbohydrate exchange. *Am J Clin Nutr* 1981; 34: 362–366.
31. Bantle JP, Wylie-Rosett J, Albright AL, et al. Nutrition recommendations and interventions for diabetes: a position statement of the American Diabetes Association. *Diabetes Care* 2008; 31(suppl 1): S61–S78.
32. Rizkalla SW, Taghrid L, Laromiguiere M, et al. Improved plasma glucose control, whole-body glucose utilization, and lipid profile on a low-glycemic index diet in type 2 diabetic men: a randomized controlled trial. *Diabetes Care* 2004; 27: 1866–1872.
33. Heilbronn LK, Noakes M, Clifton PM. The effect of high- and low-glycemic index energy restricted diets on plasma lipid and glucose profiles in type 2 diabetic subjects with varying glycemic control. *J Am Coll Nutr* 2002; 21: 120–127.
34. Ma Y, Olendzki BC, Merriam PA, et al. A randomized clinical trial comparing low-glycemic index versus ADA dietary education among individuals with type 2 diabetes. *Nutrition* 2008; 24: 45–56.
35. Jenkins DJ, Kendall CW, Keown-Eyssen G, et al. Effect of a low-glycemic index or a high-cereal fiber diet on type 2 diabetes: a randomized trial. *JAMA* 2008; 300: 2742–2753.
36. Nicholson AS, Sklar M, Barnard ND, et al. Toward improved management of NIDDM: a randomized, controlled, pilot intervention using a lowfat, vegetarian diet. *Prev Med* 1999; 29: 87–91.
37. Barnard ND, Cohen J, Jenkins DJ, et al. A low-fat vegan diet improves glycemic control and cardiovascular risk factors in a randomized clinical trial in individuals with type 2 diabetes. *Diabetes Care* 2006; 29: 1777–1783.
38. Brunerova L, Smejkalova V, Potockova J, Andel M. A comparison of the influence of a high-fat diet enriched in monounsaturated fatty acids and conventional diet on weight loss and metabolic parameters in obese non-diabetic and type 2 diabetic patients. *Diabet Med* 2007; 24: 533–540.
39. Garg A. High-monounsaturated-fat diets for patients with diabetes mellitus: a meta-analysis. *Am J Clin Nutr* 1998; 67(suppl): 577S–582S.
40. Gerhard GT, Ahinann A, Meeuws K, et al. Effects of a low-fat diet compared with these of a high monosaturated fat diet on body weight, plasma lipids, and lipoproteins and glycemic control in type 2 diabetes. *American Journal of Clinical Nutrition* 2004; 80(3): 668–673.
41. Higashi K, Shige H, Ito T, et al. Effect of a low-fat diet enriched with oleic acid on postprandial lipemia in patients with type 2 diabetes mellitus. *Lipids* 2001; 36: 1–6.
42. Gannon MC, Nuttall FQ, Lane JT, Burmeister LA. Metabolic response to cottage cheese or egg white protein, with or without glucose, in type II diabetic subjects. *Metabolism* 1992; 41: 1137–1145.
43. Gannon MC, Nuttall FQ, Grant CT, et al. Stimulation of insulin secretion by fructose ingested with protein in people with untreated type 2 diabetes. *Diabetes Care* 1998; 21: 16–22.
44. Gannon MC, Nuttall JA, Damberg G, et al. Effect of protein ingestion on the glucose appearance rate in people with type 2 diabetes. *J Clin Endocrinol Metab* 2001; 86: 1040–1047.
45. Gannon MC, Nuttall FQ, Saeed A, et al. An increase in dietary protein improves the blood glucose response in persons with type 2 diabetes. *Am J Clin Nutr* 2003; 78: 734–741.
46. Sargrad KR, Homko C, Mozzoli M, Boden G. Effect of high protein versus high carbohydrate intake on insulin sensitivity, body weight, hemoglobin A1c, and blood pressure in patients with type 2 diabetes mellitus. *J Am Diet Assoc* 2005; 105: 573–580.
47. Pomerleau J, Verdy M, Garrel DR, Nadeau MH. Effect of protein intake on glycaemic control and renal function in type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia* 1993; 36: 829–834.
48. Delahanty LM, Nathan DM. Implications of the diabetes prevention program and Look AHEAD clinical trials for lifestyle interventions. *J Am Diet Assoc* 2008; 108(suppl 1): S66–S72.

Management of Type 2 Diabetes and Gestational Diabetes in Pregnancy

Tara Boinpally and Lois Jovanovič, MD

Sansum Diabetes Research Institute, Santa Barbara, CA

ABSTRACT

Although previously thought to be predominately transient gestational diabetes, diabetes in pregnancy can be attributed more and more to type 2 diabetes today. Although all types of diabetes in pregnancy pose a threat to the health and future well-being of both the mother and child affected, undiagnosed type 2 diabetes can be significantly more devastating in complications because of effects starting from conception. This rise of type 2 diabetes thus imparts a great sense of urgency to uncover undiagnosed cases of type 2 diabetes in pregnancy and to take active measures in establishing tight glucose control. From preconception care before pregnancy to medical treatment postpartum, it is essential that immediate care be taken to help mediate the effects of diabetes in pregnancy. *Mt Sinai J Med* 76:269–280, 2009. © 2009 Mount Sinai School of Medicine

Key Words: diabetes complications, diabetes mellitus, gestational diabetes mellitus, insulin, insulin therapy, pregnancy, type 2 diabetes, macrosomia.

Over the past several years, type 2 diabetes mellitus has become a worldwide epidemic, with the number of affected individuals projected to double from the estimated 171 million in 2000 to 366 million by 2030.¹ Concomitant with this rise is the increasing prevalence of type 2 diabetes in women of child-bearing age and in pregnancy. Although true gestational diabetes occurs in the course of pregnancy, undiagnosed type 2 diabetes results in maternal hyperglycemia present from the moment of conception and therefore poses a far greater

threat to the developing fetus. As preexisting diabetes emerges as one of the most pressing and pervasive complications of pregnancy today, it is imperative that it be discovered early and teased apart from gestational diabetes in order to minimize adverse outcomes and administer more effective care.

GESTATIONAL DIABETES OVERVIEW

Definition

The clinical definition of gestational diabetes mellitus (GDM) has been changed several times within the last 2 decades, with modifications continuing today. Prior to 1991, GDM was classified as “a transient abnormality of glucose tolerance during pregnancy,”² but in 1997, the American Diabetes Association included diabetes diagnosed during pregnancy in the definition.

Today, gestational diabetes refers to “any degree of glucose intolerance with onset or first recognition during pregnancy.”³ This definition, however, fails to demarcate preexisting cases of diabetes—nearly always type 2—that are first discovered in pregnancy. Such pregnancies associated with undiagnosed type 2 diabetes can result in added risks to both the fetus and mother if not properly managed. To offer more appropriate care specialized for women with type 2 diabetes in pregnancy, Omori and Jovanovič⁴ proposed that the current definition of GDM be reconsidered to distinguish between newly occurring diabetes and long-standing diabetes. In 2 independent cohorts, Omori and Jovanovič recognized several criteria that can be used to identify women with previous type 2 diabetes in pregnancy. As presented at the 2008 International Workshop Conference on Gestational Diabetes, these criteria include the presence of acanthosis nigricans, a glucose level of 180 mg/dL after 1 hour on the oral glucose load (either the 50-g challenge or the 100-g glucose tolerance test), a systolic blood pressure >105 mm Hg, a glycosylated hemoglobin (A1C) level >5.3%, and/or a body mass

Address Correspondence to:

Lois Jovanovič

Sansum Diabetes Research
Institute

Santa Barbara, CA

Email: ljovanovic@sansum.org

index >30 —all factors strongly correlated with type 2 diabetes. After deliberation of their proposal and on the basis of these findings, the standing definition of GDM will presumably be amended to exclude pre-existing cases of type 2 diabetes first documented in pregnancy in the near future.

Causes

If we do not consider cases in which previous diabetes is misclassified as GDM, GDM is largely caused by a lack of insulin secretory capacity to compensate for the insulin resistance secondary to decreased exercise, increased food intake, increased adiposity, and the anti-insulin hormones of pregnancy. This effect on glucose metabolism is why pregnancy is often considered a diabetogenic state. In early gestation, several hormones become critical and rise dramatically in order to sustain the pregnancy, including human placental lactogen, estrogen, progesterone, prolactin, and leptin. Maternal secretion of cortisol increases as well. These same hormones also cause the mother to become slightly resistant to insulin, the hormone most essential in enabling basal glucose uptake. As the placenta continues to grow and expand during the second and third trimesters, the level of maternal and placental hormonal secretion increases, and this results in a discernible state of insulin resistance. In normal pregnancy, the maternal pancreas is able to overcome this effect by producing additional insulin. In GDM, however, the maternal pancreas is unable to sustain or achieve the appropriate output of insulin, and thus the mother's pancreatic β -cells fail to respond adequately. As a result, glucose released into the bloodstream persists and causes elevated basal and postprandial glucose levels in the mother, and this subsequently increases the level of glucose available to the developing fetus. This increased glucose available to the fetus can result in several complications in pregnancy, and thus it is imperative that blood glucose concentrations in pregnant women with diabetes be tightly controlled to minimize risks for both the mother and fetus. Usually, it is during the second trimester that GDM usually develops. In some cases, however, the condition arises in the first trimester.⁵

Although placental hormones and chronic insulin resistance coupled with decreased insulin secretion are considered the major causes of GDM, affected women also have a genetic predisposition to diabetes. Still, information assessing the contribution of genetics to the development of GDM remains sparse and is not well confirmed. Monogenic forms of diabetes, such as mitochondrial diabetes and maturity-onset diabetes of the young (which is

characterized by autosomal dominant inheritance), have been noted to contribute slightly to the risk for GDM but influence less than 5% of all GDM cases. To date, the extent to which genetics play a role in the progression of GDM remains unclear, with little evidence and few studies offering conclusive data. The genes involved in these subtypes of diabetes and GDM appear to have important effects on β -cell function, and patients often do not have evidence of chronic insulin resistance. Clinical suspicion of these subtypes is based on a lack of evidence for insulin resistance coupled with a suggestive family history. Diagnosis requires genotyping, which has recently become available for clinical practice, but there is no current evidence to suggest that genetic screening is cost-effective.

As of now, the limited data available on the genetics behind other forms of diabetes suggest modest heritability for GDM, but findings are confounded by incomplete case ascertainment. Nonetheless, the autoimmune and insulin-resistant forms of diabetes outside of pregnancy, diseases for which GDM is often a precursor, are heritable, and some contributory genetic variants have been defined. Evidence has been presented indicating that some of the variants may contribute to GDM or its physiological phenotypes (insulin resistance and β -cell dysfunction), but the studies to date are relatively small.⁶

Prevalence

Because of the different methods used for screening women, it has proven difficult to fully quantify the prevalence of GDM in the United States today. Nonetheless, using the American Diabetes Association's epidemiological data on the overall number of women with diabetes and the number of pregnancies occurring in the United States annually, we can estimate that roughly 8% to 10% of all pregnancies are complicated by diabetes. Of these pregnancies with diabetes, around 40% are cases of undiagnosed type 2 diabetes, 5% are cases of type 1 diabetes, and the remainder are cases of gestational diabetes. Those considered at low risk for GDM have a prevalence of GDM of 1.4% to 2.8%,^{7,8} whereas women at high risk appear to have rates of 3.3% to 6.1%.⁹ Race and ethnicity play considerable roles in these rates, with the prevalence of GDM highest in the Asian and Hispanic populations and lowest for Caucasian mothers.

Complications

When left untreated, diabetes in pregnancy can result in several severe fetal complications, especially within the first 6 to 7 weeks of pregnancy when

organogenesis occurs. Maternal hyperglycemia in this critical period can lead to diabetic embryopathy, the consequences of which include central nervous system deformities, musculoskeletal deformities, congenital heart diseases, and an increased risk of spontaneous abortion. Among the central nervous system malformations are neural tube defects including spina bifida, caudal agenesis, and anencephaly. Cardiac anomalies include septal defects, patent ductus arteriosus, and situs inversus defects, whereas musculoskeletal malformations include renal anomalies such as horseshoe kidney and agenesis. There is also an increased risk of macrosomia if hyperglycemia is sustained beyond the first trimester, and it can be complicated to shoulder dystocia, bone fracture, or nerve palsy during birthing.

Because of the detrimental effects of hyperglycemia in pregnancy, women with inadequately controlled preexisting diabetes face a 2- to 5-fold higher risk of having an infant with major congenital defects in comparison with women with normal glucose tolerance. Fortunately, if euglycemia is achieved prior to conception, the likelihood of adverse outcomes is equal to the risk in women without diabetes.¹⁰ If in fact preexisting diabetes is first discovered in pregnancy, however, and the potential outcomes are such that the mother chooses to continue her pregnancy, then intensive glycemic control measures must be put into place immediately.

In the later stages of pregnancy and past the period of organogenesis, hyperglycemia can result in stimulation of the fetal pancreas to secrete excessive amounts of insulin, thereby producing fetal hyperinsulinemia and increasing the child's risk of developing insulin resistance in the future. The possibility for fetal macrosomia is the complication most clearly correlated with maternal hyperglycemia. More specifically, however, it appears that the degree of transient postprandial hyperglycemia is the strongest predictor of macrosomia. These peaks of hyperglycemia have a profound effect on the child's intrauterine environment and can result in a 4-fold increased risk for macrosomia in comparison with the general population. This trend can be explained by the fact that the presumed fetal renal threshold of glucose is 110 mg/dL.¹¹ Elevations of glucose surpassing this level can result in a fetal urinary source of glucose in the amniotic fluid. After 20 weeks of gestation, the fetus begins to ingest this amniotic fluid, which takes at least 24 hours to fully turn over. Thus, even minimal peaks in maternal glucose can persist for hours in the amniotic fluid and result in overnutrition of the fetus from both oral and intravenous glucose loading.¹²

Screening and Diagnosis

Although gestational diabetes may increase thirst and/or urination in some women, the condition most commonly remains undetected until screening. Other potential symptoms include blurred vision, fatigue, frequent infections (particularly those of the bladder, vagina, and skin), nausea and vomiting, and unexplained weight loss. Such symptoms, however, are considerably nonspecific and mirror those of pregnancy, and as such, it is difficult to attribute them to diabetes alone. Hence, because of the largely asymptomatic nature of GDM, it is necessary for a woman's risk to be assessed at her first prenatal visit.¹³

Traditionally, it is recommended that physicians screen women at risk for GDM at 24 to 28 weeks' gestation. Although clinicians are advised to make decisions on a case-by-case basis regarding which patients to screen, there are several factors that indicate increased risk for developing the condition. Such factors include obesity, an age greater than 25 years, a family history of diabetes, previous gestational diabetes, and Hispanic, Native American Indian, Asian, or African American descent. For such high-risk women, screening for GDM is strongly advised, and screening may even be conducted prior to 24 weeks' gestation. If women who meet any of these criteria test negative for GDM at an early screening, they should be retested again at 24 to 28 weeks. Low-risk status can be ascribed to women who are less than 25 years of age, have normal weight prior to pregnancy, are a member of an ethnic group at low risk for GDM, have no family history of diabetes in first-degree relatives, have no history of abnormal glucose tolerance, and have no previous obstetric problems. Women with low-risk status require no glucose testing, but they must meet all the aforementioned criteria to be conclusively considered at low risk.³

Like the definition, standards for the diagnosis of GDM remain somewhat controversial and vary worldwide. Presently, GDM is screened by a 1-step or 2-step method with different diagnostic criteria of the outcomes outlined by the American Diabetes Association, American College of Obstetrics and Gynecologists, or World Health Organization. The 1-step method, most commonly employed outside the United States, consists of an oral glucose tolerance test (OGTT) in which a 75- or 100-g oral glucose load is administered following a fasting state without a previous plasma or serum screening. In the United States, the 2-step method is customarily used. This method involves the administration of an initial 50-g, 1-hour glucose challenge test followed by an OGTT to confirm the diagnosis of GDM in patients

who have an initial result of ≥ 130 mg/dL in the first glucose challenge test.^{3,14} With either the 1-step or 2-step approach, the diagnosis of GDM is based on an OGTT. However, soon there will be a worldwide consensus to use only the 1-step, 75-g oral glucose load.

Recently, however, there have been great efforts to standardize the screening methods and test criteria used to diagnose GDM, and consensus may soon be reached now that results from the Hyperglycemia and Adverse Pregnancy Outcome study have been published. This 7-year international study was designed to both establish criteria for GDM to prevent fetal macrosomia and standardize glucose thresholds for the relationship between maternal hyperglycemia and adverse perinatal outcomes. After using the 75-g OGTT approach and correlating primary outcomes with fasting, 1-hour, and 2-hour time points, the researchers discovered that the fasting plasma glucose concentration seems sufficient in comparison with the other tested time points for identifying pregnant women who are at risk for adverse outcomes.¹⁵ As such, it is alleged that a fasting OGTT using a 75-g glucose load will soon be internationally adopted on the basis of these new data. GDM will be classified in individuals with a fasting glucose level >95 mg/dL, and more conclusive testing will be employed for individuals in the range of 75 to 95 mg/dL. Women with glucose levels <75 mg/dL will be considered to have normal glucose tolerance in pregnancy and will not be characterized as having GDM.

TREATMENT AND MANAGEMENT IN PREGNANCY

Preconception

In women with preexisting diabetes, preconception planning is fundamental to having a successful and healthy pregnancy. Because of the grave effects of hyperglycemia in the child's period of early development and organogenesis, it is imperative that women with diabetes prior to pregnancy be especially vigilant about their blood glucose concentrations. A1C levels, which indicate an individual's average blood glucose levels over the past 2 to 3 months, are excellent markers to use in offering recommendations regarding readiness for pregnancy. Although the standard target for a nonpregnant diabetic patient is an A1C level of $<6.5\%$, research indicates that A1C levels 6.0% decrease the risk for malformations to a level comparable to the risk in nondiabetic pregnancies. Prior to pregnancy, a target goal of <150 mg/dL

is suitable for the 1-hour postmeal glucose test, and in both the preconception and pregnancy periods, premeal and fasting glucose target levels should be <90 mg/dL. Thus, to meet these more intensive targets, women with preexisting cases of type 2 diabetes should discontinue use of oral antihyperglycemic agents (OAHAs) and begin insulin therapy, whereas women with type 1 diabetes and those already using insulin should take care to adjust their treatment appropriately. Blood pressure, kidney function, and retinopathy should also be assessed and properly treated in all diabetic women in order to help prevent worsening during pregnancy. As with all other women, a full obstetric and gynecological examination is considered an integral component of preconception care, along with the adoption of a healthy prepregnancy lifestyle incorporating good nutrition and exercise. Special care should also be taken for women who experienced gestational diabetes in any previous pregnancy. Those women with a fasting plasma glucose reading in excess of 95 mg/dL post-conception should be evaluated and treated as if they have preexisting diabetes.

Although most complications associated with diabetes in pregnancy have to do with the child's health, women with preexisting type 1 or type 2 diabetes have added risks that make preconception planning of further importance. Diabetic retinopathy progression may be triggered by a rapid normalization of blood glucose concentration occurring during or prior to pregnancy. This risk can be minimized by the gradual achievement of greater metabolic control in the preconception period, and tests should be done to assess a woman's retinopathy status prior to pregnancy. If retinopathy is found, it is advised that pregnancy be delayed until her retinal status is treated and stabilized. If retinopathy is absent, then she should be notified about the 10% chance of developing retinopathy in pregnancy and have at least 1 full retinal examination by an ophthalmologist every trimester.

Preexisting hypertension must also be addressed in order to prevent worsening of other diabetic comorbidities in pregnancy, such as nephropathy or retinopathy. Women on drug therapy for hypertension may need to begin alternate treatment regimens to make sure that their medications will not adversely affect the pregnancy. Before pregnancy, those with uncontrolled blood pressure could start short-term use of angiotensin-converting enzyme (ACE) inhibitors to lower blood pressure and improve proteinuria in tandem with an effective contraceptive method to prevent conception. Once blood pressure levels are under control, the ACE inhibitor and contraceptive should be discontinued

and replaced with alternate blood pressure therapies documented to be safe in pregnancy so that the woman can conceive. Renal insufficiency is also associated with preexisting diabetes and can negatively affect fetal development, and thus renal function should be evaluated prior to pregnancy. This baseline assessment should include a 24-hour urine collection to determine total protein excretion, creatinine clearance, and microalbuminuria. Women with renal disease are at increased risk for fetal mortality and morbidity, and heavy proteinuria also increases the chance of poor pregnancy outcomes. In severe cases, pregnancy may be discouraged, or alternatives may be discussed.¹⁶

Medical Nutritional Therapy

Fundamental to treating insulin resistance and diabetes in pregnancy are lifestyle interventions that incorporate diabetes-specialized nutritional therapy and exercise. Many women with diabetes in pregnancy, with the exception of mothers with type 1 diabetes, can achieve good glucose control via medical nutritional therapy (MNT). The optimum nutritional management would be a diet that meets the appropriate caloric and nutritional needs of the pregnant woman while maintaining euglycemia throughout the course of the pregnancy. Especially important in this regimen is the prevention of postprandial hyperglycemia, as such glucose level peaks are associated with an increased risk of fetal macrosomia and birth trauma.

The American Diabetes Association's recent statement on nutrition recommendations and interventions for diabetes advises that all women with diabetes in pregnancy receive MNT at diagnosis. MNT involves a carbohydrate-controlled meal plan with specific dietary recommendations determined and regularly modified by individual assessment. The regimen relies on the monitoring of blood glucose concentrations, urine ketones, appetite, and weight gain. The American Diabetes Association position statement also recommends that women with diabetes in pregnancy obtain adequate caloric intake for appropriate weight gain. Yet, although the American Diabetes Association does not recommend weight loss, carbohydrate restriction and moderate weight loss are considered suitable for some overweight and obese women with type 2 diabetes or GDM. The concern about hypocaloric diets in obese women stems from the associated risk of ketonemia and ketonuria from ketoacidosis or starvation ketosis. Moderate caloric restriction (reduction by 30% of estimated energy needs), however, is deemed a potential approach to reduce maternal weight gain

and improve glycemic control without subsequent ketonemia.¹⁷

In addition to minimizing macrosomia, MNT in pregnancy should also focus on preventing maternal postprandial hyperglycemia in order to lessen diabetic fetopathy from fetal exposure to excess glucose. This peak postprandial response can best be avoided with a carbohydrate-restricted diet. Although there have not yet been any trials conducted to assess the optimal diet for either lean or obese women with GDM, it appears that the euglycemic diet consisting of 30 kcal/kg of present pregnant weight for normal weight women, 24 kcal/kg for overweight women, and 12 kcal/kg for morbidly obese women satisfies the needs of pregnancy without leading to excessive weight gain or hyperglycemia. In comparison with the diet proposed by the American Diabetes Association, which advocates a 55% carbohydrate diet, the euglycemic diet decreases carbohydrate content to 40% and incorporates more fat. The diet also recommends the following overall caloric distribution: 12.5% at breakfast, 28% at both lunch and dinner, and the remainder divided among 3 snacks. There is also a limitation of carbohydrates to 33% at breakfast, as insulin resistance appears to be greatest in the morning. The euglycemic diet is based on studies illustrating the direct relationship between the postprandial glucose response and the carbohydrate content of the diet and has thus far proven successful in helping women achieve substantial glycemic control.¹⁷

Exercise can also play an important role in lifestyle interventions employed to minimize the complications of diabetes in pregnancy. It is recommended that all pregnant women capable of exercising take part in at least 30 minutes of planned physical activity each day. Ways for women with GDM to meet this goal include doing arm exercises before mealtime while seated and walking briskly. Studies show that adopting a regular exercise regimen is particularly beneficial for women with GDM, as it helps lower fasting and postprandial glucose concentrations.¹⁷

Achieving Euglycemia

Despite efforts to maintain euglycemia with diet and exercise alone, many women with gestational diabetes discover that hyperglycemia persists, necessitating supplemental pharmacological therapy. Considered by some to be a stringent guideline, maintaining an A1C range of 4% to 6% in pregnancy greatly improves the likelihood of a healthy outcome, and thus it is critical that every pregnant woman's A1C level be as close to these values as possible. Although many different therapies today are being considered as avenues of treatment for GDM, it is still likely that

an insulin regimen will need to be started in order to maintain this target A1C range. Additionally, because the 1-hour postprandial glucose peak appears to play the largest role in the development of fetal macrosomia, insulin therapy should be specially designed to treat this peak as well. Proper insulin therapy thus combines different types of insulin to maintain both the basal glucose level and the postprandial peak.¹⁶

To mediate the postprandial peak, a rapid-acting insulin should be used immediately before each meal. Because regular human insulin tends to aggregate, it is advised that rapid-acting insulin analogues be used in its place. As of now, there are 2 main types of rapid-acting insulin analogues that have been investigated and proven safe in pregnancy: insulin lispro and insulin aspart. Compared to regular human insulin, insulin lispro therapy lowers A1C in women with GDM¹⁸ and has also been shown to improve the postprandial glucose peak in comparison with regular insulin.¹⁹ Moreover, insulin lispro is unlikely to cross the placenta,²⁰ and this makes it a safe choice in pregnancy. The other insulin analogue considered safe to use for the rapid-acting mealtime bolus is insulin aspart. According to research in pregnant woman with type 1 diabetes, and in comparison with human insulin, the use of insulin aspart leads to reduced nocturnal hypoglycemia and greater postprandial control,²⁰ along with a decrease in preterm deliveries and fetal loss.²¹ Insulin glulisine is also a short-acting insulin analogue, but it is under pregnancy category C and is not recommended for use in pregnancy because of a lack of clinical data to show its safety. In general, rapid-acting insulin analogues are a key component of an insulin regimen in pregnancy as they allow for a decrease in hypoglycemic events and an increase in postprandial glucose control when taken just prior to a meal.

The second component of insulin therapy is an insulin that acts throughout the day as the basal insulin release. The only effective choice as of now is neutral protamine Hagedorn (NPH) insulin, which has a peak effect 4 to 12 hours after administration and a duration of action of up to 26 hours. Because of the timing of its peak effect, it is necessary for a pregnant woman to take 3 injections of NPH (1 every 8 hours) in order to maintain her glucose control with basal insulin. Although long-acting insulin analogues exist, such as insulin glargine and insulin detemir, they are labeled as pregnancy category C because of a lack of clinical trials done in pregnancy to establish their safety and efficacy. Additionally, case study reports show that mothers on glargine do not achieve the same tight control as women on NPH. However, although insulin detemir is also not recommended for pregnancy, there is currently

Table 1. *Insulin Requirement for Pregnancy. The constant "k" is used to derive the total daily dosage ($k \times w$) in kg.*

Gestational week	0–12	13–28	29–34	35–40
<i>k</i>	0.7	0.8	0.9	1.0

Total units of insulin per day = Weight (kg) $\times$ *k* (U/kg/day) = *X* (U/day)

a large multinational clinical trial studying the safety and efficacy of insulin detemir in pregnancies complicated by type 1 diabetes. As an alternative to NPH, the rapid-acting insulin analogues aspart and lispro can be used safely in a continuous insulin infusion pump to provide the basal needs.²³

In order for insulin therapy to be effective in preventing hyperglycemia and postprandial peaks, special care must be taken to calculate a woman's individual insulin requirements and to adjust dosing levels throughout the course of her pregnancy. This repeated fine-tuning is made in accordance with a woman's week of gestation and her response to her personal regimen. On the basis of studies on well-controlled diabetic women and extensive observations in the field of gestational diabetes, Jovanović²⁴ developed an algorithm to determine baseline insulin requirements and dosages. The initial step is to calculate the total amount of insulin required daily (Table 1).

Incorporated into this calculation is the constant *k*, which changes according to the woman's gestational week and represents the total daily units of insulin per kilogram of body weight. From weeks 0 to 12, this *k* value is 0.7 U/kg/day; from weeks 13 to 28, this value is 0.8 U/kg/day; from weeks 29 to 34, this value is 0.9 U/kg/day; and from weeks 35 to 40, this value is 1.0 U/kg/day. Multiplying the appropriate constant *k* by a woman's weight in kilograms gives the total units of insulin to be prescribed per day:

Total units of insulin per day =

$$\text{Weight (kg)} \times k \text{ (U/kg/day)} = X \text{ (U/day)}$$

For example, the insulin requirement of a 50-kg woman at 15 weeks' gestation would be $50.0 \text{ kg} \times 0.8 \text{ U/kg/day} = 40 \text{ U/day}$. The numerical answer of this calculation is thus the total amount of insulin required over 24 hours. This value should subsequently be split evenly into the basal insulin requirement needed during the fasting state and the bolus insulin requirement necessary for meals:

Basal insulin requirement = $0.50 \times$ Total daily units

Bolus insulin requirement = $0.50 \times$ Total daily units

The total basal insulin (the preferred choice being NPH) is then divided into 3 equal injections to be given every 8 hours, or at 8 am, 4 pm, and 12 am. The remaining bolus insulin, which should be a rapid-acting insulin, must also be divided between mealtimes. However, because meals can vary in size and in the carbohydrate content to which they affect the blood glucose concentration, bolus insulin can initially be separated into thirds for breakfast, lunch, and dinner but can then be adjusted over time on the basis of premeal and postmeal blood glucose levels.

Just as important as choosing the correct type of insulin for therapy is establishing a schedule of blood glucose self-monitoring and insulin administration. Women with GDM should test their blood concentrations up to 8 times a day, incorporating a reading before every mealtime and at 1 hour post-meal, at bedtime, and in the middle of the night. Information from these values can be used by her physician to fine-tune an insulin regimen accordingly. For all pregnant women with diabetes, including those with GDM, fasting glucose target levels should be <90 mg/dL, whereas 1-hour post-meal values should be <120 mg/dL. Although such goals may seem stringent, following these guidelines results in a substantial decrease in the risk of negative outcomes associated with diabetic pregnancies.¹⁶ Intrinsic to reaching these goals and tight control is the fine adjustment of a woman's insulin regimen in response to fluctuations in glucose levels. After baseline insulin dosages are determined, the amount of insulin administered should be titrated in accordance with the results of periodic glucose monitoring. Boluses before mealtime should be adjusted on the basis of 1-hour postprandial blood sugar readings and increased or decreased on the following day by 2 units of insulin if the blood sugar levels are <110 mg/dL or >121 mg/dL, respectively (Table 2). For example, if a woman's blood sugar 1 hour after lunch is 80 mg/dL, then the next day's prelunch rapid insulin bolus should be decreased by 2 units.

NPH insulin boluses should also be corrected regularly, as gauged by fasting glucose measures as well as readings taken throughout the day (Table 3).

Table 2. Next day's insulin dosage based on today's blood glucose concentration on hour after meals.

1-Hour Postprandial Blood Sugar	Next Day's Insulin Adjustment for Corresponding Premeal Bolus*
<110 mg dL	Decrease insulin by 2 units
>121 mg dL	Increase insulin by 2 units

*The 1-hour postprandial lunch level leads to the adjustment for the next day's prelunch bolus.

Table 3. Adjustment of the next day's insulin dosage based on today's pre-prandial glucose concentration.

Blood Sugar Reading	Adjustment for NPH Insulin Bolus
Before dinner	Next day's morning NPH dosage
<60 mg/dL	Decrease insulin by 2 units.
>91 mg/dL	Increase insulin by 2 units.
Before bedtime	Next day's dinner NPH dosage
<60 mg/dL	Decrease insulin by 2 units.
>91 mg/dL	Increase insulin by 2 units.
Before breakfast	Same day's bedtime NPH dosage
<60 mg/dL	Decrease insulin by 2 units.
>91 mg/dL	Take 3 am reading at night: If >70 mg/dL, decrease next bedtime NPH by 2 units. If >91 mg/dL, call a physician for a rapid insulin scale equal to the prelunch rapid insulin scale.

Abbreviation: NPH, neutral protamine Hagedorn.

Because intermediate-acting insulin, such as NPH, covers basal insulin requirements for approximately 8 hours, it is recommended that multiple blood sugar readings be taken, including prebreakfast, predinner, and bedtime readings. If the predinner blood sugar is <60 mg/dL, then the next day's morning NPH dosage should be decreased by 2 units so that this low reading does not occur again. This holds true for a prebreakfast glucose reading of <60 mg/dL, which would necessitate a change in the same day's bedtime NPH, as well as the reading before bedtime, which would require a decrease in the predinner NPH. Additionally, upward adjustments must be made if blood sugars run too high. For example, if yesterday's predinner reading was >91 mg/dL, then today's morning NPH must be increased by 2 units. This modification also applies to yesterday's blood sugar reading before bedtime and the current day's dinner NPH dosage. However, if today's prebreakfast reading is >91 mg/dL, then a 3 am blood sugar reading should be taken. If it is >70 mg/dL (regardless of today's prebreakfast blood sugar), then bedtime NPH that evening should be increased by 2 units. If it is >91 mg/dL, then the doctor should be called for a 3 am rapid insulin scale equal to the prelunch rapid scale.²⁴

Use of Oral Antihyperglycemic Agents (OAHAs)

Although insulin therapy with injections and glucose monitoring may at first appear challenging, women often find the routine more and more manageable as the pregnancy progresses. Additionally, it is presently the only recommended medicinal treatment for gestational diabetes in the United States. New

findings do indicate, however, that the use of OAHAs to supplement insulin or act in place of insulin may be effective in treating women with only GDM but OAHAs would not be appropriate for women with preexisting diabetes mellitus. Nevertheless, their safety and value throughout the entire term of pregnancy have not been fully ascertained.

Today, glyburide and metformin are the most common drugs used for treating diabetes. Glyburide is a sulfonylurea that induces pancreatic beta-cells to release insulin, and metformin is a biguanide used to enhance insulin's effect in the body. Both glyburide and metformin cross the placenta to some extent, although glyburide does so to a lesser extent. As of now, studies show that glyburide is not currently effective in preventing macrosomia, neonatal hypoglycemia (or neonatal hyperinsulinemia), or other neonatal complications and thus appears less effective than insulin therapy.^{24,25} Conversely, metformin appears to be as safe and effective as insulin during the second and third trimesters of pregnancy with respect to reducing perinatal complications.²⁶ Traditionally, metformin has been used in nondiabetic women with polycystic ovary syndrome to reduce spontaneous abortions and increase fertility, and it does not appear to increase the risk of congenital defects or induce hypoglycemia in the mother when used in the first trimester.^{28,29} Research does indicate, however, that many women with diabetes who tried using metformin alone were unable to reach their target glucose control goals and had to supplement it with insulin. Additionally, use of these OAHAs does not appear to mediate the postprandial glucose peak that results in excess glucose persisting in the amniotic fluid. Therefore, although there is much promise for the future with respect to using OAHAs for women with purely transient gestational diabetes, insulin therapy remains the best standard of care for all women with preexisting diabetes.

Other Diabetic Drugs in Pregnancy

It is also important to evaluate the use of hypertensive drugs during pregnancy; these drugs are commonly used by women with type 2 diabetes mellitus. As ACE inhibitors and angiogenesis receptor blocker drugs are contraindicated in pregnancy and may cause congenital malformations, alternate methods of hypertension control must be used.

Within the first 12 weeks of pregnancy, calcium channel blockers should also be avoided. Currently, methyldopa and hydralazine are the only blood pressure-lowering medications considered appropriate for use in pregnancy, and they are considered the

safest of blood pressure-lowering medications for use during the first trimester of pregnancy. Because these are short acting drugs, their dosage should be spaced out through the day. Other more intensive blood pressure-lowering drugs, such as nifedipine and labetalol, should be used only in situations in which the potential benefits outweigh the risk of use.¹⁶

POSTPARTUM TREATMENT AND MANAGEMENT

Implications of Gestational Diabetes Postpartum

Although women with gestational diabetes (preexisting cases of diabetes are thus excluded) regain normal glucose tolerance postpartum, they are at high risk for experiencing GDM in subsequent pregnancies. It is important that a woman at high risk for recurring GDM speak with her physician if she intends to have another child in order to ensure normal fetal development.

Women with previous GDM are also at risk for developing diabetes mellitus later in life, with studies indicating a 20% to 50% chance of developing it within the next 5 to 10 years after pregnancy.^{30,31} Because of this high probability, it is important that follow-up care start soon after the delivery as a precautionary measure and to ensure prompt diagnosis if in fact diabetes develops.

After delivery, women with GDM may return to normal glucose tolerance, continue with existing beta-cell deterioration and develop type 2 diabetes, or, in rare cases, develop type 1 diabetes. Unfortunately, there are currently no treatments available to prevent type 1 diabetes mellitus, but new immunomodulatory therapies in development may offer some help in the future. With respect to subsequent type 2 diabetes, on the other hand, there are many protective measures that can be taken. In most cases, gestational diabetes presents as a result of chronic insulin resistance, and thus the progression to type 2 diabetes mellitus is very probable without proper care. Within the 6 to 12 weeks following the delivery and for every year thereafter, glucose monitoring incorporating an OGTT should be initiated to check for diabetes progression and impaired glucose tolerance. Blood pressure testing and lipid testing are vital as well, as many women will have characteristics of metabolic syndrome.⁶

Most precautions that help lower the risk for developing type 2 diabetes mellitus for women with previous GDM revolve around lifestyle modifications that help lessen insulin resistance. Results from

the Diabetes Prevention Program showed that those randomly assigned to an intensive lifestyle intervention were able to reduce the risk of type 2 diabetes by 58%, thus confirming that exercise and weight loss are very effective in delaying or preventing the development of type 2 diabetes.³² It is advised that women who have experienced GDM try to get at least 30 minutes of physical activity 5 days a week and return to their prepregnancy weight within a year of the delivery. If still overweight after this initial weight loss, women are encouraged to try to lose at least 5% to 7% of their body weight.³²

For some women with advanced insulin resistance and impaired glucose tolerance, however, lifestyle modifications are not enough to delay or prevent diabetes postpartum. In such situations, medications used to treat insulin resistance may be of some benefit. The results of TRIPOD (Troglitazone in Prevention of Diabetes) and its follow-up PIPOD (Pioglitazone in Prevention of Diabetes) demonstrated that Hispanic women with previous GDM who were treated with the insulin-sensitizing thiazolidinedione drug troglitazone had a greater than 50% reduction in the occurrence of type 2 diabetes mellitus over a treatment period of 30 months. Additionally, those who participated in the subsequent PIPOD study using pioglitazone over another 3 years of treatment had beta-cell function stabilization. Thus, it appears that women with previous GDM who face a progressive loss of insulin secretion could be greatly helped by the implementation of medications that reduce insulin resistance, such as thiazolidinedione drugs or even metformin.^{22,27,33,34}

Importance of Breast Feeding

Whether a mother has had gestational diabetes or not, breastfeeding offers significant health benefits for the child, ranging from increased protection against developing type 1 diabetes to helping avoid childhood obesity. The advantages of breastfeeding are often attributed to the colostrum of breast milk, which contains a great variety of protective factors and antibodies. As such, breastfeeding is considered very valuable for the growing infant, especially with respect to the baby's ability to build tolerance or reject certain foods and dietary components.³⁵ This plays an especially important role when we consider the prevention of type 1 diabetes mellitus. Early weaning from breast milk within the first 2 to 3 months of life has been shown to increase the risk of developing type 1 diabetes mellitus. This may be due to a loss of natural protective agents or the possible introduction of cow's milk in formula, as many studies have revealed a link between exposure to cow's milk

within the first 3 months of life (or sometimes even later in life) and the increased prevalence of type 1 diabetes.³⁶⁻³⁸

This association between type 1 diabetes and cow's milk is thought to be a result of the body's immune response to biological compounds found within the bovine milk. One theory is that the ingested cow's milk may contain bovine insulin, for which the body would begin making antibodies. These immunoglobulins appear to cross-react with human insulin, as bovine insulin differs from human insulin by only 3 amino acids, making it immunogenic.³⁶⁻³⁸ Another plausible explanation is that antibodies created against cow's milk proteins of bovine albumin, beta-lactoglobulin, and beta-casein may mistakenly attack the child's own beta-cells because of amino acid sequence similarity between the milk proteins, beta-cell surface receptors, and glucose transporters.³⁹⁻⁴¹ Such molecular mimicry could thus potentially lead to autoimmune destruction. Thus, it is likely that an early introduction to cow's milk has particularly strong implications for children who have a genetic predisposition to type 1 diabetes. As such, special care should be taken in determining whether to feed with breast milk or cow's milk and when to introduce cow's milk. Yet, although much evidence exists, there is still speculation due to conflicting studies, and thus the association with cow's milk and increased risk of type 1 diabetes remains unclear.⁴²

It is also true that infants of type 2 mothers may fare better when they are breastfed rather than bottlefed, as trends show that women who are overweight tend to overfeed their babies. In fact, the Endocrine Society considers it part of clinical guidelines of practice to encourage breastfeeding infants for at least 6 months in order to help prevent obesity.⁴³

One note of caution, however, is that a woman with diabetes must check glucose levels prior to breastfeeding her child. Current research suggests that excess maternal glucose overflows into the breast milk past a maternal glucose threshold of 150 mg/dL. Glucose concentrations surpassing the threshold will lead to an excess of sugar in the breast milk, resulting in reactive hypoglycemia and increasing the infant's food-seeking behavior. This behavior then increases the tendency for the mother to overfeed an irritable baby. In instances in which the mother's sugar level exceeds 150 mg/dL, it is better that the child receive formula or previously stored euglycemic breast milk.⁴⁴

Table 4. *Key Points.*

The definition of gestational diabetes will likely be amended to demarcate preexisting type 2 diabetes from true gestational diabetes according to the criteria of a glucose level of 180 mg/dL after 1 hour on the oral glucose load (either the 50-g challenge or the 100-g glucose tolerance test), the presence of acanthosis nigricans, a systolic blood pressure >105 mm Hg, a glycosylated hemoglobin level >5.3%, and/or a body mass index >30.
When left untreated, diabetes in pregnancy can result in diabetic embryopathy, central nervous system deformities, musculoskeletal deformities, congenital heart diseases, an increased risk of spontaneous abortion, and an increased risk of macrosomia.
Standards for screening methods and test criteria used to diagnose gestational diabetes mellitus may soon be set on the basis of data from the recently published Hyperglycemia and Adverse Pregnancy Outcome study.
Fundamental to treating insulin resistance and diabetes in pregnancy are lifestyle interventions that incorporate diabetes-specialized nutritional therapy and exercise.
Especially important is the prevention of postprandial hyperglycemia, as such glucose level peaks are associated with increased risks of fetal macrosomia and birth trauma.
Supplemental medicinal therapy is often required to maintain euglycemia. Maintaining a glycosylated hemoglobin range of 4% to 6% in pregnancy greatly improves the likelihood of a healthy outcome, and the described insulin regimen has been designed to achieve this.
After delivery, women with gestational diabetes mellitus may or may not return to normal glucose tolerance. It is imperative that all women follow the appropriate postpartum guidelines, including breast feeding and insulin monitoring.

CONCLUSION

As the prevalence of type 2 diabetes mellitus continues to rise at an epidemic rate, it is now more important than ever that preexisting cases be diagnosed prior to pregnancy in order to ensure healthy fetal development. It is our belief that the most effective way to improve the fate of the next generation is to employ a more proactive diabetes screening process for women within their child-bearing years. This preconception care would necessitate that all healthcare practitioners who routinely work with women, including obstetricians, family physicians, dietitians, and gynecologists, make it standard care to test for diabetes in all women of child-bearing age. Two aspects in particular would be central to this preconception counseling: a discussion of plans for pregnancy and a check for diabetes as a preventative measure. Thus, appointments made for pap smears, for pelvic examinations, or even for obtaining birth control prescriptions should also include this preconception counseling, regardless of the initial reason for the visit (Table 4).

Thorough screenings conducted in this manner would likely provide the most comprehensive approach to identifying women with type 2 diabetes prior to any subsequent pregnancy. The resulting earlier detections would then lead to a decrease in complications related to uncontrolled maternal hyperglycemia within the period of organogenesis. Despite the marked improvements in patient care, it is understandable that such a program would take time to be integrated. Inexorably requiring great changes

in the current opinions of diabetes management in pregnancy, this method necessitates a redefinition of gestational diabetes to distinguish women with type 2 diabetes as a unique diagnostic category during pregnancy as well as a consensus on the most appropriate screening methods. It is our hope, however, that as our understanding of diabetes and pregnancy continues to expand, such changes will be adopted in an effort to provide the best possible care for all women in pregnancy and the future generations that they bear.^{22–34}

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Wild S, Roglic G, Green A, et al. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004; 27: 1047–1053.
2. O'Sullivan JB. Gestational diabetes and its significance in early diabetes. In: Camerini-Davalos R, Cole HS, eds. *Early Diabetes*. New York, NY: Academic 1970; 339–344.
3. US Preventive Services Task Force. Screening for gestational diabetes mellitus: US Preventive Services Task Force recommendation statement. *Annals* 2008; 148: 759–765.
4. Omori Y, Jovanović L. Proposal for the reconsideration of the definition of gestational diabetes [letter]. *Diabetes Care* 2005; 28: 2592–2593.
5. Carr DB, Gabbe S. Gestational diabetes: detection, management and implications. *Clin Diabetes* 1998; 16: 4–11.

6. Metzger BE, Buchanan TA, Coustan DR, et al. Summary and recommendations of the Fifth International Workshop-Conference on Gestational Diabetes Mellitus [published correction appears in *Diabetes Care* 2007; 30:3154.]. *Diabetes Care* 2007; 30(suppl 2): S251–S260.
7. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2006; 29(suppl 1): S43–S48.
8. Jovanović L, Pettitt DJ. Gestational diabetes mellitus. *JAMA* 2001; 286: 2516–2518.
9. Marguette GP, Klein VR, Niebyl JR. Efficacy of screening for gestational diabetes. *Am J Perinatol* 1985; 2: 7–9.
10. Jovanović L, ed. *Medical Management of Pregnancy Complicated by Diabetes*. 3rd ed. Alexandria, VA: American Diabetes Association 2000.
11. Ostertag S, Jovanović L, Lewis B, Auld P. Insulin infusion therapy for low birth weight infants. *Pediatrics* 1986; 78: 625–630.
12. Jovanović-Peterson L, Peterson CM, Reed GF, et al. Maternal postprandial glucose levels and infant birth weight: the Diabetes in Early Pregnancy Study. *Am J Obstet Gynecol* 1991; 164: 103–111.
13. Ross G. Gestational diabetes. *Aust Fam Physician* 2006; 35: 392–396.
14. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2008; 31: S55–S60.
15. HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008; 358: 1991–2002.
16. Jovanović L. The importance of preconception care in women with preexisting diabetes. *Curr Diab Pract* 2006; 6(1): 1–4.
17. Jovanović L. Achieving euglycaemia in women with gestational diabetes mellitus: current options for screening, diagnosis and treatment. *Drugs* 2004; 64(13): 1401–1417.
18. Bhattacharyya A, Brown S, Hughes S, et al. Insulin lispro and regular insulin in pregnancy. *QJM* 2001; 94: 255–260.
19. Meccaci F, Carignani L, Cioni R, et al. Maternal metabolic control and perinatal outcome in women with gestational diabetes treated with regular or lispro insulin: comparison with non-pregnant diabetic women. *Eur J Obstet Gynecol Reprod Biol* 2003; 11: 19–24.
20. Boskovic R, Feig DS, Derewlany L, et al. Transfer of insulin lispro across the human placenta. *Diabetes Care* 2003; 26: 1390–1394.
21. Matheisen ER, Kinsley B, Amiel SA, et al. Maternal glycemic control and hypoglycemia in type 1 diabetic pregnancy: a randomized trial of insulin aspart versus human insulin in 322 pregnant women. *Diabetes Care* 2007; 30: 771–776.
22. Hod M, Damm P, Kaaja R, et al. Fetal and perinatal outcomes in type 1 diabetes pregnancy: a randomized study comparing insulin aspart with human insulin in 322 subjects. *Am J Obstet Gynecol* 2007; 198: 186.e1–186.e7.
23. Jovanović L, Kitzmiller J. Insulin therapy in pregnancy. In: Hod M, Jovanović L, Di Renzo GC, et al., eds. *Textbook of Diabetes in Pregnancy*. London, United Kingdom: Informa Healthcare 2008; 205–216.
24. Jovanović L. Insulin therapy and algorithms for treating type 1 diabetes mellitus. Available at: <http://www.diabetesroundtable.com/courses/optinsulin/algor1.asp>.
25. Langer O, Conway DL, Berkus MD, et al. A comparison of glyburide and insulin in women with gestational diabetes mellitus. *N Engl J Med* 2000; 343: 1134–1138.
26. Langer O, Yogeve Y, Xenakis EM, Rosenn B. Insulin and glyburide therapy: dosage, severity level of gestational diabetes, and pregnancy outcome. *Am J Obstet Gynecol* 2005; 192: 134–139.
27. Rowan JA, Hague WM, Gao W, et al. Metformin versus insulin for the treatment of gestational diabetes. *N Engl J Med* 2008; 358: 2003–2015.
28. Coetzee EJ, Jackson WP. Oral hypoglycaemics in the first trimester and fetal outcome. *S Afr Med J* 1984; 65: 635.
29. Gilbert C, Valois M, Koren G. Pregnancy outcome after first-trimester exposure to metformin: a meta-analysis. *Fertil Steril* 2006; 86: 658–663.
30. National Institute of Diabetes and Digestive and Kidney Diseases. *National diabetes statistics fact sheet: general information and national estimates on diabetes in the United States, 2005*. Bethesda, MD: US Department of Health and Human Services; 2005.
31. American Diabetes Association. Gestational diabetes mellitus (position statement). *Diabetes Care* 2004; 27(suppl 1): S88–S90.
32. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002; 346: 393–403.
33. Azen SP, Peters RK, Berkowitz K, et al. TRIPOD (Troglitazone in the Prevention of Diabetes): a randomized, placebo-controlled trial of troglitazone in women with prior gestational diabetes mellitus. *Control Clin Trials* 1998; 19: 217–231.
34. Buchanan TA, Xiang AH, Kjos SL, et al. Diabetes rates and b-cell function in the Pioglitazone in Prevention of Diabetes (PIPOD) study. In: Program and abstracts of the 65th Scientific Sessions of the American Diabetes Association; June 10–14, 2005; San Diego, CA. Abstract 157-OR.
35. Paronen J, Knip M, Savilahti E, et al. Cow's milk formula feeding induces primary immunization to insulin in infants at genetic risk for type 1 diabetes. *Diabetes* 2000; 49: 1657–1665.
36. Verge CF, Gianani R, Yu L, et al. Late progression to diabetes and evidence for chronic beta cell autoimmunity in identical twins of patients with type 1 diabetes. *Diabetes* 1995; 44: 1176–1179.
37. Vaarala O, Paronen J, Otonkoski T, Akerblom HK. Cow milk feeding induces antibodies to insulin in children—a link between cow milk and insulin-dependent diabetes mellitus? *Scand J Immunol* 1998; 47: 131–135.
38. Vaarala O, Knip M, Paronen J, et al. Cow milk feeding induces primary immunization to insulin in infants at genetic risk for type 1 diabetes. *Diabetes* 1999; 48: 1389–1394.
39. Strobel S, Mowat A. Immune responses to dietary antigens: oral tolerance. *Immunol Today* 1998; 19: 173–181.

40. Karjalainen J, Martin JM, Knip M, et al. A bovine albumin peptide as a possible trigger of insulin-dependent diabetes mellitus. *N Engl J Med* 1992; 327: 302–307.
41. Vaarala O, Klemetti P, Savilahti E, et al. Cellular immune response to cow's milk B-lactoglobulin in patients with newly diagnosed IDDM. *Diabetes* 1996; 45: 178–182.
42. Virtanen SM, Läärä E, Hyppönen E, et al. Cow's milk consumption, HLA-DQB1 genotype, and type 1 diabetes mellitus. A nested case-control study of siblings of children with diabetes. *Diabetes* 2000; 49: 912–917.
43. August GP, Caprio S, Fennoy I, et al. Prevention and treatment of pediatric obesity: an Endocrine Society clinical practice guideline based on expert opinion. *J Clin Endocrinol Metab* 2008; 93: 4576–4599.
44. Jovanović-Peterson L, Fuhrmann K, Hedden K, et al. Maternal milk and plasma glucose and insulin levels: studies in normal and diabetic subjects. *J Am Coll Nutr* 1989; 8: 125–131.

Role of Surgery in Management of Type II Diabetes

Daniel M. Herron, MD, FACS, and Winnie Tong, MD

Section of Bariatric Surgery, Division of Laparoscopic Surgery, Department of Surgery, Mount Sinai School of Medicine, New York, NY

ABSTRACT

Diabetes is one of the most common diseases in the United States, with 1.5 million new cases reported in 2005 and a projected 48.3 million diabetics in the year 2050. Despite the disease's remarkably high prevalence and the substantial efforts focused on pharmacological treatment of the disease, current dietary, behavioral, and medical therapies have achieved only limited success. For over 2 decades, bariatric surgeons have recognized that gastric bypass surgery causes a remarkable, durable improvement in type 2 diabetes mellitus. Remarkably, this normalization of glucose metabolism occurs within several weeks of surgery, long before substantial weight loss takes place. As early as in 1987, data have been available demonstrating that 83% of patients with diabetes and 99% of those with impaired glucose tolerance become euglycemic after gastric bypass surgery. Despite the availability of this effective treatment for diabetes, the primary therapy for type 2 diabetes mellitus has remained medical. In this article, we review some of the hypotheses about how bariatric operations affect glucose metabolism and diabetes as well as the research that has been done to elucidate these mechanisms. Additionally, we discuss how surgery may one day represent an important management option for type 2 diabetes mellitus. *Mt Sinai J Med* 76:281–293, 2009. © 2009 Mount Sinai School of Medicine

Key Words: bariatric surgery, diabetes, metabolic surgery.

Address Correspondence to:

Daniel M. Herron, MD, FACS

Section of Bariatric Surgery

Division of Laparoscopic
Surgery

Department of Surgery

Mount Sinai School of Medicine

New York, NY

Email: daniel.herron@mountsinai.org

Diabetes is one of the most common diseases in the United States, with 1.5 million new cases reported in 2005 and a projected 48.3 million diabetics in the year 2050.¹ Despite this remarkably high prevalence and the large amount of effort focused on pharmacological treatment of the disease, current dietary, behavioral, and medical therapies have achieved only limited success. Large prospective trials have demonstrated that hemoglobin A1C (HgA1C) levels may remain elevated (between 7.5 and 8.5) even with intensive medical therapy.²

The glycemic control obtained from medical therapy for type 2 diabetes mellitus (T2DM) is effective in some patients, inadequate in others, and always dependent on continued administration of oral or injected medication. This stands in marked contrast to what is seen after bariatric surgery. For over 2 decades, bariatric surgeons have recognized that gastric bypass surgery causes a remarkable, durable improvement in T2DM. Remarkably, this normalization of glucose metabolism occurs within several weeks of surgery, long before substantial weight loss takes place.

In an early landmark report from 1987, Walter Pories³ presented his data on 141 patients with diabetes or impaired glucose tolerance who had undergone gastric bypass at East Carolina University School of Medicine. Of these 141 individuals, all but 2 became euglycemic within 10 days after surgery. He followed up his initial series 8 years later in a publication entitled "Who Would Have Thought It? An Operation Proves To Be the Most Effective Therapy for Adult-Onset Diabetes Mellitus."⁴ In this series, with a much longer follow-up (mean, 7.6 years; total, 1254 patient years), 83% of gastric bypass patients with preoperative T2DM and 99% of those with impaired glucose tolerance ultimately maintained normal levels of plasma glucose, HgA1C, and insulin.

Despite the availability of this remarkably effective treatment, the primary therapy for T2DM has remained medical. Current guidelines for the

treatment of T2DM from the American Diabetes Association do not even mention surgery in the treatment algorithm.⁵ However, as the prevalence of both obesity and T2DM increases at unprecedented rates, the need for a cure, as opposed to a treatment for this chronic disease, has never been greater. At present, roughly one-third of adults in the United States are obese.⁶ Over 24 million Americans have been diagnosed with diabetes, and 800,000 cases are newly diagnosed each year. Total costs incurred from diabetes in 2007 amounted to \$174 billion. As of 2006, diabetes was the seventh leading cause of death in the United States.⁷

In this article, we review some of the hypotheses about how bariatric operations affect glucose metabolism and diabetes as well as the research that has been done to elucidate these mechanisms. Additionally, we discuss how surgery may one day fit within the overall management strategy for T2DM.

CURRENT BARIATRIC OPERATIONS AND THEIR MECHANISM OF ACTION

A number of different bariatric procedures are commonly performed in the United States. For the past decade, the Roux-en-Y gastric bypass (RYGB; Figure 1) has been the most common operation, with an estimated 103,000 procedures performed in the United States in 2003.⁸ Initially performed through a laparotomy—the traditional open approach—the operation is now typically performed with a laparoscopic technique through 5 to 6 small incisions in the abdominal wall. Multiple firings of a surgical stapler are used to divide the stomach into 2 discrete segments: the upper stomach pouch, which is approximately 15 to 30 mL or less in volume, and the lower, bypassed gastric remnant. The jejunum is surgically divided 50 to 100 cm beyond its origin at the ligament of Treitz and reconnected in a Y-shaped fashion such that one arm of the Y (the Roux limb or alimentary limb) drains the small gastric pouch, whereas the other provides egress for the gastric secretions, bile, and pancreatic juices originating from the gastric remnant and duodenum.⁹

The laparoscopic adjustable gastric band (Figure 2), popularly called the lap band, is currently the second most common bariatric procedure, and it is rapidly increasing in popularity, partially because of an aggressive advertising campaign by its manufacturer (Allergan, Irvine, CA). In this procedure, an adjustable plastic and silicone ring is placed around the upper stomach just beneath the gastroesophageal junction. A subcutaneous access port

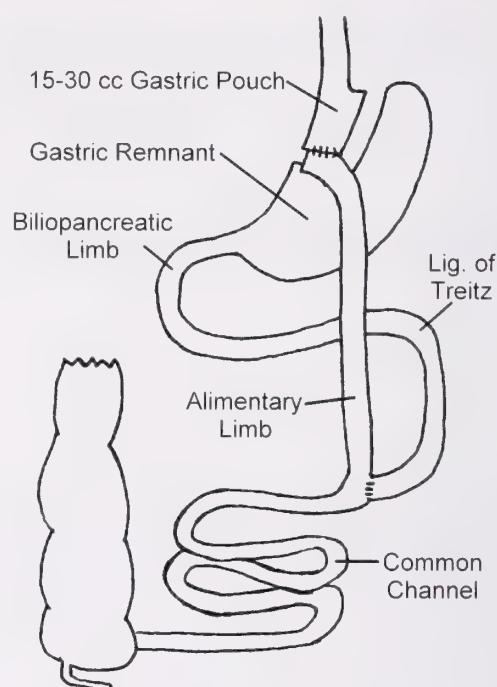


Fig 1. Roux-en-Y gastric bypass. A 15- to 30-mL pouch is created at the upper stomach, draining into the Roux or alimentary limb. Reprinted with permission. Copyright 2008, Daniel M. Herron.

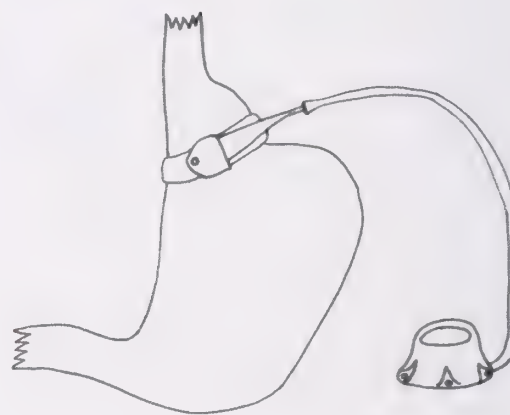


Fig 2. Laparoscopic adjustable gastric band. The inflatable band is placed just inferior to the esophagogastric junction. Reprinted with permission. Copyright 2008, Daniel M. Herron.

allows the degree of band constriction to be adjusted by the injection or withdrawal of a saline solution from the system. Unlike the gastric bypass, the gastric band is a purely restrictive operation and does not bypass any segment of the gastrointestinal tract. Although the risk of mortality and major morbidity is less with this operation than with gastric bypass, the amount of excess weight loss obtained is also less.¹⁰

Gaining in popularity is the sleeve gastrectomy (Figure 3). This procedure, typically performed



Fig 3. Sleeve gastrectomy. The greater curvature of the stomach is resected, and this leaves a tubularized stomach with a volume of 150 to 200 mL. **Abbreviation:** B-P, biliopancreatic. Reprinted with permission. Copyright 2008, Daniel M. Herron.

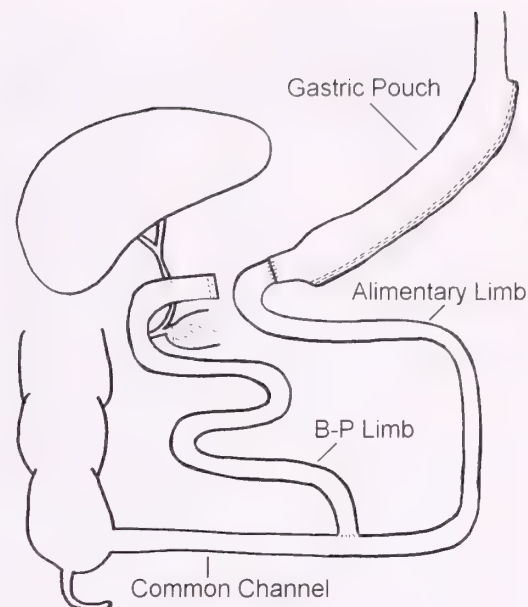


Fig 4. Biliopancreatic diversion with a duodenal switch. After a sleeve gastrectomy, the first portion of the duodenum is connected to the distal 250 cm of the small bowel. The remaining biliopancreatic limb is anastomosed 100 cm proximal to the ileocecal junction. Reprinted with permission. Copyright 2008, Daniel M. Herron.

laparoscopically, involves resection of the greater curvature, or left side, of the stomach. The stomach is stapled over a sizing tube ranging from 11 to 20 mm in diameter, and this creates a stomach roughly the size and shape of a banana. By reducing the volume of the stomach, the sleeve gastrectomy creates a substantial restrictive effect. The operation also removes most of the fundus of the stomach, where ghrelin, a hunger-promoting hormone, is secreted; this may result in systemic effects beyond restriction, which are discussed in depth later.¹¹

The most malabsorptive bariatric operation is the biliopancreatic diversion with a duodenal switch (BPD-DS; Figure 4). This operation consists of a sleeve gastrectomy combined with a severely malabsorptive intestinal bypass. The first portion of the duodenum is surgically divided and connected to the last 250 cm of ileum. The remaining intestine, through which the bile and pancreatic secretions pass, is anastomosed 100 cm from the ileum. This creates an anatomy that is less restrictive than a gastric bypass or gastric band but substantially more malabsorptive.⁹

Traditionally, several mechanisms of action have been invoked to describe how bariatric operations result in weight loss. All bariatric operations provide some degree of gastric restriction, in that the functional volume of the stomach is reduced. This is accomplished by the reduction of the size of the stomach, as in gastric bypass and sleeve gastrectomy, or by the creation of a small upper stomach pouch with a constricted outlet, as in the adjustable gastric band. Most bariatric surgeons feel that the gastric bypass additionally provides a small amount of malabsorption because the ingested food is separated

from digestive enzymes (gastric acid, bile, pancreatic amylase, and lipase) for part of its journey through the gastrointestinal tract. Recently, investigators have become increasingly aware that hormonal effects may have an important impact on postsurgical weight loss.¹¹ Bariatric operations result in decreased stimulation—or outright elimination in the case of the sleeve gastrectomy—of the gastric fundus, the area of the stomach in which ghrelin, a recognized hunger-promoting hormone, is produced.

CLINICAL STUDIES EVALUATING THE RESOLUTION OF DIABETES AFTER BARIATRIC SURGERY

Many investigators have described the beneficial effects of bariatric surgery on T2DM mellitus; these are summarized in Table 1. In one of the earliest reports, the Pories group at East Carolina University reported the complete normalization of blood glucose in 121 of 146 patients (82.9%) with T2DM who underwent gastric bypass.⁴ Additionally, 150 of the 152 patients in that study with impaired glucose tolerance (98.7%) were cured after surgery.

These statistics are supported by other researchers in more recent reports. A University of Pittsburgh study from 2003 reported on the follow-up of 191

Table 1. Results for Diabetes Control After Bariatric Surgery.

Reference	Number of Patients	Mean Age (Years)	EW Change (%)	Type of Surgery	Follow-Up (Years)	Diabetic Outcome	Study Type
18 34	1025 413	39 48.7	NA -58	RYGB GB	7-10 2	94% cured 80% cured	RC RC
113 114	88 1250	47.5 NA	-51 -56	GB GB	0.5-1.6 5	65.3% cured 64% cured, 26% improved, 10% unchanged	RC Mixed
27 115	500 126	NA 49.5	-38 -46	GB SG	1 1	Improved 85% cured, 11% improved	RC RC
116	91	49	SG, -63; RYGB, -66	SG, RYGB	1	SG, 84% cured; RYGB, 84.6% cured	RC
26 117 118	443 312 342	42 43 43	NA NA -58	BPD BPD RYGB	20 10 0-14	91.2% cured 97% cured 85% cured, 10% improved	RC RC PC
119 12	143 1160	42.9 48	NA -60.1	GB RYGB	3 0.5-4.5	83% cured 83% cured, 17% improved	PC PC
16 29	117 41	44 44.6	NA NA	RYGB SG	1 1.9	74% cured, 26% improved 76% cured, 15% improved	PC PC
24	22,094	39	RYGB, -61; VBG, -68; BPD, -70; GB, -50	RYGB, VBG, BPD, GB	> 0.08	76% cured, 8% improved, 0.7% deteriorated	Meta-analysis
120	S, 851; C, 853	46	RYGB, -25; VBG, -16; GB, -13; C, +1.6	RYGB, GB, VBG	10	S, 36% cured; C, 13% cured	Cohort
121	S, 17; C, 20	54.5	S, -13; C, +2	GB	4	S, 45% cured; C, 4% cured	Cohort
122	S, 40; C, 40	40	S, -87; C, -21	GB	2	Surgery better than controls	RCT
123	100	37	GB, -54.9; VBG, -70	GB, VBG	2	GB, 8% cured; VBG, 13% cured	RCT

Abbreviations: BPD, biliopancreatic diversion; C, control group; EW, excess body weight; GB, gastric banding; NA, not applicable; PC, prospective cohort; RC, retrospective cohort; RCT, randomized control trial; RYGB, Roux-en-Y gastric bypass; S, surgery group; SG, sleeve gastrectomy; VBG, vertical banded gastroplasty.

patients with T2DM who had undergone laparoscopic gastric bypass at that institution over a 5-year period.¹² Fasting plasma glucose and HgA1C levels normalized in 83% of patients and improved substantially in the remaining 17%.

One of the highest-quality ongoing studies of bariatric surgery to date has been performed by the Swedish Obese Subjects Scientific Group. These investigators are following 4047 subjects who were treated for obesity with either surgery or medical management.¹³ In one report from this group, 72% of surgically treated patients demonstrated recovery from T2DM 2 years after surgery, whereas only 21% of control patients did. The unadjusted prevalence of diabetes among controls increased from 7.8% to 24.9% after 8 years. However, the prevalence of diabetes in the surgical group remained relatively stable (10.8% at inclusion versus 10.5% at 8 years, $P < 0.0001$).¹⁴ In another study from the same group, 767 obese patients who underwent surgery for weight loss were compared to 712 matched, conventionally treated patients. The incidence of diabetes was significantly lower in the surgical group than in the control group at 2 years (0.2% versus 6.3%, respectively) and 10 years (7% versus 24.9%, respectively).^{13,15}

In general, RYGB is associated with a reduction in diabetic medication requirements in 80% to 98% of patients at up to 14 years of follow-up.^{12,16–18} Clinical improvement in diabetes appears to be most pronounced after RYGB in patients with milder and shorter duration of the disease.^{12,16} Patients with less central obesity, as measured by a smaller waist circumference, also have greater improvement in their diabetes postoperatively.¹⁶ On the other hand, patients who fail to resolve their diabetes after RYGB tend to be older, suffer from staple line breakdown, or have a more prolonged preoperative disease course in comparison with their counterparts in whom diabetes was cured.^{4,18,19}

Studies comparing RYGB to medical therapy suggest that surgery is more effective than medical treatment alone in improving diabetes over the long term. McDonald *et al.*¹⁷ compared 154 morbidly obese T2DM patients who underwent RYGB to 78 controls who did not. At a mean follow-up of 6.2 years for the surgical group, the antiglycemic medication requirement fell from 31% to 9% ($P = 0.0001$). In contrast, at 9 years' follow-up, 87% of the control group required antiglycemic medication. This represented a significant disease progression from a baseline requirement of 56% in the control group ($P = 0.0003$).

Another study evaluated 109 morbidly obese patients who underwent open RYGB and compared

them to 27 control patients who elected not to undergo surgery.²⁰ At a mean follow-up period of 5.8 years, only 1 of 109 surgical patients (0.9%) developed diabetes versus 6 controls (22%). The single surgical patient who remained diabetic lost weight initially but subsequently regained weight.

BPD-DS is a substantially more involved operation than RYGB, with a higher risk of morbidity and mortality; for this reason, it is less commonly performed. However, the existing clinical evidence suggests that it is extremely effective in ameliorating diabetes, with normalization of glucose and insulin levels in over 90% of patients up to 20 years postoperatively.^{12,21–24} A report by Scopinaro *et al.*²³ on 195 obese diabetic patients who were followed for 10 years after biliopancreatic diversion (BPD) found that 98% of patients experienced resolution of their diabetes. Only 3 subjects had fasting serum glucose levels in the diabetic range. Notably, no patients, including those with incomplete recovery or mild relapse of their diabetes, required diabetic treatment during the study period.

In another study, 25 BPD patients were followed after surgery, and their insulin resistance and acute insulin response—2 of the earliest defects in the course of T2DM—were measured.²⁵ Preoperatively, 12 patients had T2DM, 5 had impaired glucose tolerance, and 8 had normal glucose tolerance. All patients had reduced insulin sensitivity at baseline. Acute insulin response, which would normally increase after an intravenous glucose tolerance test in healthy subjects, was absent in the diabetic group. Acute insulin response was reduced in the impaired glucose tolerance group and appropriately increased in the normal group preoperatively. Twelve months after surgery, all patients achieved complete resolution of T2DM with normalization of their acute insulin response and insulin sensitivity. These improvements occurred even though the patients remained classified as obese despite substantial weight loss.

A more recent study by Scopinaro *et al.*²⁶ reported that the likelihood of diabetes remission after BPD was related to the duration of diabetes and specific preoperative antidiabetic treatment findings that were similar to those after RYGB. Diabetes resolution occurred in a higher proportion of patients after BPD versus those undergoing restrictive procedures or RYGB.^{12,17,18,27} Despite these remarkable rates of diabetes improvement with BPD-DS, the high complication rate and technical difficulty of the operation have continued to limit its popularity in the United States.²⁸

When compared to bypass procedures, adjustable gastric banding (GB) has been shown to be somewhat less effective in achieving diabetes resolution.

Several studies have suggested that fewer than 50% of patients treated with banding achieve remission of their diabetes.^{13,24} Dixon and O'Brien²⁷ followed 50 patients with T2DM for 1 year after gastric band placement. Of patients who relied on oral medication preoperatively, 64% normalized their fasting serum glucose, insulin, and glycosylated hemoglobin while off medication. Diabetic control improved in another 26% of gastric band patients who were able to decrease the dosages of their diabetes medications. This study was limited by the relative short duration of follow-up. Another study reported glycemic control to be less pronounced in obese diabetic patients after banding in comparison with RYGB.¹³

Laparoscopic sleeve gastrectomy has shown promising results with respect to diabetes improvement. In 17 superobese diabetic patients with a mean body mass index (BMI) of 57.3 kg/m², diabetes was cured in 76.9% and improved in 15.4% of subjects with up to 18 months of follow-up.^{26,29} Other reports suggest resolution of diabetes in 84% to 95% of super-obese patient 1 year after sleeve gastrectomy.^{30,31}

Vertical banded gastroplasty (VBG), popularly known in the 1980s as stomach stapling, has been shown to have poor long-term weight loss results and has fallen out of favor in the United States.³³ However, even this operation has been shown to be beneficial for diabetes control in morbidly obese patients. Studies on VBG have demonstrated resolution of diabetes in 71% to 83% of patients with up to 4 years of postoperative follow-up.²⁴

A meta-analysis of 361 bariatric surgery studies with 85,048 patients was published by Buchwald *et al.* in 2004.²⁴ The analysis showed an overall diabetes remission rate of 76.8% with improvement in 86% of patients. Diabetic remission occurred in 98.9% of patients after BPD or BPD-DS, in 83.7% after RYGB, in 71.6% after VBG, and in 47.9% after adjustable gastric banding.

Recently, a remarkable study was published by the Centre for Obesity Research and Education at Monash University in Melbourne, Australia.³³ The group performed a prospective, randomized trial in which 60 patients with a BMI between 30 and 40 kg/m² and T2DM were assigned to either surgery (adjustable gastric band) or medical therapy. The primary outcome variable was remission of diabetes, as measured by a fasting glucose level < 126 mg/dL and HgA1C < 6.2%. Of the 53 patients who completed the 2-year study, remission of diabetes was observed in 22 patients (73%) in the surgery group but in only 4 patients (13%) in the medically treated group.

PROPOSED MECHANISMS BY WHICH BARIATRIC SURGERY IMPROVES DIABETES

The pathophysiology by which bariatric surgery achieves improvement of diabetes is not well understood but is most likely related to some combination of weight loss, decreased caloric intake, intestinal malabsorption, hormonal changes, and rearrangement of the gastrointestinal anatomy. Each of these possible mechanisms is discussed here in greater depth.

Role of Weight Loss

Given that T2DM is closely associated with obesity, weight loss was initially believed to be the mechanism primarily responsible for the resolution of diabetes after bariatric surgery. The best model for assessing this conjecture is the adjustable gastric band, as this operation does not change gastrointestinal anatomy beyond constriction of the upper stomach. In Dixon *et al.*'s study,³² the rate of diabetes remission after GB corresponded to the degree of weight loss. Additionally, there are data suggesting that resolution of T2DM after GB is more prominent as patients lose more weight; Ponce's group³⁴ reported better weight loss and T2DM resolution 2 years after the operation than after the first year.

In contrast, net weight loss does not appear to be a primary mechanism for diabetes resolution after RYGB and BPD-DS. Most studies have demonstrated that resolution of T2DM occurs within days after surgery, before any significant weight loss has been achieved.^{19,25} Patients with diabetes or impaired glucose tolerance experience improvement in insulin resistance, as measured by intravenous glucose tolerance tests and homeostasis model assessment, as early as 6 days after the operation. This antidiabetic effect is sustained up to at least 1 year after surgery and is not related to weight loss, which proceeds at a much slower rate.³⁵ Another report showed that all morbidly obese patients who normalized their glucose levels 1 month after BPD were still more than 80% over their excess body weight.³⁶

Interestingly, another study demonstrated that plasma levels of fasting plasma glucose, insulin, and leptin are lower in postoperative RYGB patients maintaining stable weight in comparison with nonsurgical controls matched for weight, age, and body fat percentage.⁵⁷

Still, the relationship between weight loss after RYGB and the magnitude of diabetes resolution remains somewhat controversial. Although some

studies have shown that the percentage of excess weight loss correlates with diabetes resolution, others have not.^{12,16,18}

Role of Decreased Caloric Intake

Decreased caloric intake is another postsurgical mechanism that could be responsible for better short-term glycemic control. Pories *et al.*⁴ reported on a patient whose bariatric surgery was canceled because of a full stomach. Despite failing to undergo surgery, the patient was maintained on a strict postoperative diet identical to that of the postoperative gastric bypass patients. During this time, his glucose and insulin levels normalized. Perhaps not surprisingly, the patient was unable to maintain such an extreme diet for a prolonged period, and his diabetes recurred.

On the other hand, a number of human and animal studies have suggested that reduced caloric intake is not a major factor in diabetes improvement after bariatric surgery. A recent rat study by Rubino *et al.*³⁸ showed that diabetic animals who underwent duodenal jejunal bypass (DJB) demonstrated improved glycemic control with respect to sham-operated animals that had the same food intake and body weight. Similarly, human patients whose diabetes did not improve despite a very low calorie diet later experienced diabetes remission after RYGB.³⁹ Furthermore, patients only temporarily limited their food intake after BPD, but their blood glucose remained under control even after they recovered full eating capacity.³⁶ Gastropasty, an operation that significantly reduces gastric volume and oral intake, is not as effective in improving hyperglycemia and hyperinsulinemia as RYGB.⁴⁰

Role of Intestinal Malabsorption

Existing data suggest that malabsorption may play a role in T2DM resolution after BPD-DS but not other procedures. BPD-DS has been found in some series to be even better than gastric bypass in resolving T2DM.²⁸ This operation induces a remarkable reduction in intestinal fat absorption and availability of free fatty acids, selectively depleting intramyocellular lipids in skeletal muscle with subsequent restoration of glucose utilization as the primary energy source and improved insulin resistance. These physiological changes occur despite the presence of excess total body fat in the early postoperative months after the surgery.⁴¹ In contrast, intestinal malabsorption is not evident after standard gastric bypass.^{11,42–44} In the animal model, diabetic rats did not show reductions in carbohydrate or lipid absorption after DJB but did demonstrate improvements in glucose tolerance.³⁸

Hormonal Effects

The observation that diabetes resolution after bariatric surgery occurs before major weight loss suggests that glycemic control may be hormonally mediated. This is supported by multiple studies showing that hormonal changes after bypass procedures, such as BPD and RYGB, are greater than those following restrictive procedures such as VBG and adjustable GB.^{15,45}

Incretins are peptides that are secreted after a meal and serve to enhance insulin secretion. T2DM is characterized by multiple functional deficiencies in various organs, including attenuated glucose transport into the skeletal muscle, pancreatic beta-cell impairment, increased hepatic glucose output, and an altered effect of incretins. Two particular incretins, glucagon-like peptide 1 (GLP-1) and gastric inhibitory polypeptide (GIP), have garnered particular attention because of their antidiabetic effect after surgery.

GLP-1 and its response following bariatric surgery have been extensively examined over the past 2 decades. GLP-1, also known as enteroglucagon, is secreted by the L-cells of the terminal ileum in response to a mixed meal.^{47,48} Under physiological conditions, GLP-1 acts to suppress glucagons, reduce glucose production by the liver, increase insulin production, slow gastric emptying, and enhance satiety.^{47,49–53} At least in rodents, GLP induces beta-cell neogenesis and proliferation and inhibits apoptosis.⁵⁴

Case reports of beta-cell hypertrophy and hyperplasia with hyperinsulinemic hypoglycemia following RYGB and BPD suggest that increased levels of GLP or other beta-cell trophic factors could contribute to excess beta-cell proliferation after these operations.⁵⁵ Several studies have shown a significantly attenuated postprandial GLP-1 level in T2DM patients compared to matched nondiabetic controls.^{56–59} An infusion of GLP-1 in diabetic dogs has been shown to improve glucose control.⁶⁰ Similarly, a continuous infusion of GLP-1 in diabetic patients lowered fasting and mean plasma glucose levels, decreased glycosylated hemoglobin concentrations, and enhanced insulin sensitivity and beta-cell function.^{51,61} Evidence is accumulating that GLP-1 is increased after various bariatric operations, including BPD, jejunoileal bypass, RYGB, VBG, and ileal transposition.^{53,62–69} However, some authors are less convinced of the role of GLP-1 in glycemic control; one study revealed an increase in GLP-1 after RYGB in patients with normal or impaired glucose tolerance but not in those with diabetes.^{70,71}

In contrast to GLP-1, there are fewer studies on GIP, and their results are less conclusive. GIP is an incretin secreted by K-cells in the proximal small intestine that stimulates insulin

release and synthesis.^{72,73} Notably, GIP and GLP-1 both have a synergistic insulintropic effect during hyperglycemia.⁷⁴ GIP levels are normal during fasting and after meal ingestion in diabetics, but the effect of administered GIP on insulin secretion is blunted.^{56,75,76} There are conflicting data on the effect of bariatric surgery on GIP levels. GIP has been shown to decrease or increase after jejunoileal bypass.^{62,64,77,78} After RYGB, GIP was shown to decrease, but it returned to basal concentrations more expediently than in matched controls.^{66,79-81}

Apart from incretins, a host of other hormones originating from the absorptive process that are altered after bariatric surgery have been investigated to determine if they play an integral role in the maintenance of glucose homeostasis. To date, however, no such role has been clearly established, and the results from these studies are inconclusive.

Peptide YY (PYY) is a hormone secreted by endocrine L-cells in the distal ileum in response to meals. PYY attenuates appetite and has been investigated, in the form of a nasal spray, as a weight-loss medication.^{82,83} Postprandial PYY levels have been found to be elevated after gastric bypass procedures in the early postoperative period in both humans and rodents.^{46,67,70,84,85} This elevation persists for at least 2 years in humans, with a direct relation to satiety and weight loss.⁶⁸ However, the effect on insulin in humans is unclear; one study failed to reveal any changes in insulin or glucagon concentrations after PYY administration.⁸⁶

Acylation-stimulating protein (ASP) is another hormone for which there are very limited data. ASP increases glucose uptake and fatty acid esterification in a synergistic effect with insulin. ASP has been shown to decrease⁸⁷ after RYGB, GB, and BPD, but its role in diabetes control has yet to be determined.

Leptin is an adipocyte-derived hormone. The mechanisms by which leptin regulates glucose homeostasis are not well elucidated. Upon stimulation of hypothalamic leptin receptors, neuropeptide Y levels decrease, suppressing appetite and increasing the metabolic rate. The relationship between leptin and insulin resistance is not well understood. In *in vitro* studies, leptin appears to suppress insulin secretion⁸⁸⁻⁹⁰ by reducing glucose transport and cellular adenosine triphosphate levels and by activating adenosine triphosphate-sensitive potassium channels leading to cellular hyperpolarization.^{89,90-92} Other authors have suggested that a drop in insulin resistance leads to a decrease in leptin.⁹³

It is also controversial whether body fat composition is the sole factor that regulates leptin. Multiple studies have demonstrated that leptin concentration rises after fat deposition and decreases

after RYGB, GB, and BPD.^{37,80,94} Leptin levels have been shown to decrease proportionally to fat mass at 2 years after GB⁹⁵ and to BMI at 1 year after RYGB.^{96,97} However, other investigators have not found any correlation with leptin or BMI after bariatric surgery.⁹³

Evidence is accumulating on the effect of adiponectin on glycemic control. Adiponectin is an adipose tissue-specific factor that stimulates glucose utilization and fatty acid oxidation in skeletal muscle.⁹⁸ Adiponectin is decreased in obesity⁹⁹ and correlates inversely with weight, BMI, and waist-to-hip ratio.^{100,101} There is speculation that long-term improvement in diabetes control may be mediated by a loss of fat mass and resultant changes in adipocytokine levels of the adioinsular axis.

Adiponectin has been shown to increase after RYGB, GB, and BPD.^{46,87,96,100} The administration of adiponectin in rats improves glycemic control by reducing hepatic gluconeogenesis and increasing fatty acid oxidation in muscle, thus improving insulin sensitivity.¹⁰² However, high levels of adiponectin appear to be a risk factor for the development of diabetes.¹⁰³

Ghrelin, an appetite-stimulating hormone produced in the fundus of the stomach, has generated substantial controversy with respect to its role in diabetes control. Ghrelin levels increase 3 to 4 times a day and appear to be related to appetite stimulation around mealtime. Cummings *et al.*¹¹ showed that the diurnal variation in ghrelin is suppressed after gastric bypass and that levels drop substantially after surgery. Ghrelin has been shown to have both stimulatory and inhibitory effects on insulin secretion. One study showed that diversion of food from the gastric antrum and duodenum initially increased ghrelin secretion but later suppressed it. Studies showed enhanced postprandial suppression of ghrelin after RYGB⁸¹ but did not demonstrate any change in its level after total parenteral nutrition.¹¹ Other studies have demonstrated that the mean fasting plasma ghrelin concentrations before and after GB surgery do not differ substantially.¹⁰⁴

The hormones described here and others may act in concert with one another or synergistically with incretins to affect glucose homeostasis. The existing literature implicating these hormones in T2DM after surgery is inadequate and inconclusive. Carefully designed studies are needed to improve our understanding of diabetes and determine the hormonal changes responsible for improved glycemic control after bariatric surgery.

Rearrangement of Gastrointestinal Anatomy

In the gastric bypass operation, food passes from the esophagus into a small gastric pouch and then directly into the jejunum. This anatomic change results in two distinct effects. First, the stomach, duodenum, and proximal jejunum are bypassed and thus excluded from the food path. Second, ingested food is shunted more rapidly to the distal small bowel. These two effects have been invoked to explain the nearly immediate impact of gastric bypass on glucose homeostasis, and they are called the foregut hypothesis and the hindgut hypothesis.

Foregut Hypothesis

The foregut hypothesis states that the elimination of food exposure to the hormonally active antrum, duodenum, and proximal jejunum decreases the secretion of factors that induce insulin resistance and T2DM.^{105,106} This hypothesis is supported by the observation that bypass procedures (RYGB and BPD-DS) result in better control of diabetes than pure restrictive procedures (GB).⁴⁵

In an attempt to clarify the effect of foregut exclusion on diabetes, Rubino³⁸ performed either duodenal jejunal bypass (DJB) surgery or gastrojejunostomy (GJ) in Goto Kakizaki type 2 diabetic (GK) rats. In the former group, food was completely diverted from the duodenum and proximal jejunum; in the latter, it was not. DJB rats had markedly better glucose tolerance than sham-operated control or GJ rats 10 days after the operation. Diabetes recurred when the duodenal passage was restored in the DJB rats.

It should be noted that both operations created a shortcut for ingested nutrients to pass to the hindgut more rapidly. However, the GJ rats, who still had food passing through their foregut, did not enjoy an improvement in their diabetes. When the GJ rats underwent a second operation to exclude the proximal intestine by the conversion of the GJ to a DJB, there was an amelioration of glucose intolerance in these animals. Furthermore, normoglycemic Wistar rats that underwent DJB had worsened glucose tolerance in comparison with sham-operated controls.⁴⁸ This study suggested that the presence of mediating factors from the foregut could contribute to the pathogenesis of T2DM and that bypassing this portion of the small intestine could ameliorate the disease. This bypass procedure appears to be disruptive to glucose control in nondiabetic animals.

Other studies looking at hormonal changes after DJB surgery in animals have demonstrated mixed results. In one study, DJB rats displayed higher GLP-1 levels in comparison with sham-operated controls

4 weeks after the operation.¹⁰⁷ In contrast, another study found no change in GLP-1 or GIP levels in DJB rats compared to control rats 1 week after surgery. DJB rats exhibited decreased fasting glucose and glucose levels after an oral glucose tolerance test. There was significant attenuation in leptin levels after glucose challenge in DJB rats, and this suggested a role for leptin in mediating glucose homeostasis in the jejunal bypass model.¹⁰⁸

Hindgut Hypothesis

In contrast, the hindgut hypothesis suggests that a signal promoting diabetes control may be enhanced by the expedited presentation of undigested enteric contents to the mid-jejunum (in RYGB) or distal ileum (in BPD-DS).^{53,109–111} Several groups have investigated this theory by performing ileal transposition, an operation in which a segment of the distal ileum is taken out and reconnected more proximally in the jejunum.

In one study, obese nondiabetic rats fed a high-fat diet demonstrated increased insulin sensitivity after ileal transposition in comparison with controls with intestinal transections and reanastomosis without transposition. Ileal transposition rats did not display any difference in glucose tolerance with respect to controls. Additionally, experimental rats had greater weight loss and enhanced GLP-1 and PYY postoperatively in comparison with controls.⁸⁵

Patriti *et al.*¹¹² performed ileal transposition in GK rats. Ileal transposition improved glucose tolerance, insulin sensitivity, and acute insulin response in GK rats up to 5 months after surgery in comparison with sham-operated and nonoperated animals. Investigators noted restoration of islet cell morphology in the ileal transposition rats resembling that of euglycemic rats. Ileal transposition rats also displayed more sustained GLP-1 secretion in comparison with sham-operated rats.

A recent study attempted to distinguish the effect of foregut bypass from the effect of rapid hindgut exposure on T2DM control by comparing GK rats that underwent ileal transposition to those that underwent DJ bypass.¹⁰⁷ DJ-bypass rats exhibited similarly improved glucose tolerance in comparison with ileal transposition rats 4 weeks after the operation. However, DJ-bypass rats displayed lower GLP-1 levels after the oral glucose tolerance test in comparison with the ileal transposition group at 1 week after the operation. GLP-1 levels were higher in both operated groups with respect to sham controls by 4 weeks. These results suggest that both foregut exclusion and expedited hindgut exposure improve glucose tolerance, but the mechanisms may differ. GLP-1 may play more of a role in mediating

diabetes control in rapid hindgut exposure, whereas other factors may mediate the antidiabetic effects in foregut exclusion procedures.

CONCLUSION

The use of bariatric surgery to achieve and maintain weight loss and ameliorate or resolve diabetes is generating great interest in light of the current epidemic of obesity and diabetes worldwide. Traditional medical therapies for the treatment of T2DM are evolving and improving along with our understanding of the disease's underlying pathophysiology. At present, bariatric surgery is recognized as the most effective therapy for treating severe obesity and its associated metabolic abnormalities in patients who have been unsuccessful with dietary, behavioral, and medical interventions. However, its use as an intervention for T2DM is not yet recognized in patients with a BMI below 35 kg/m².

Certainly, there remains a tremendous amount that we do not know about surgical intervention in the management of diabetes. At present, the literature is lacking with respect to prospective randomized studies and a thorough evaluation of long-term outcomes after surgery.

Nonetheless, we feel it is likely that emerging data regarding the beneficial metabolic effects of bariatric surgery will support its increasing use in the treatment of diabetes as well as obesity. As we learn more about glucose homeostasis and the mechanisms of action of bariatric surgery, we are hopeful that surgery may one day be a valid option for primary treatment of T2DM.

DISCLOSURES

Potential conflict of interest: Disclosures for Tong: None. Disclosures for Herron:

Covidien – Educational grant
Ethicon – Educational grant
Gore – Consultant
Hourglass Technology – Consultant

REFERENCES

1. Narayan KMV, Boyle JP, Geiss LS, et al. Impact of recent increase in incidence on future diabetes burden. *Diabetes Care* 2006; 29: 2114–2116.
2. Abaira C, Henderson WG, Colwell JA, et al. Response to intensive therapy steps and to glipizide dose in combination with insulin in type 2 diabetes. VA feasibility study on glycemic control and complications (VA CSDM). *Diabetes Care* 1998; 21: 574–579.
3. Pories WJ, Caro JF, Flickinger EG, et al. The control of diabetes mellitus (NIDDM) in the morbidly obese with the Greenville gastric bypass. *Ann Surg* 1987; 206: 316–323.
4. Pories WJ, Swanson MS, MacDonald KG, et al. Who would have thought it? An operation proves to be the most effective therapy for adult-onset diabetes mellitus. *Ann Surg* 1995; 222: 339–350.
5. American Diabetes Association. Standards of medical care in diabetes—2008. *Diabetes Care* 2008; 31(suppl 1): S12–S54.
6. Ogden CL, Carroll MD, Curtin LR, et al. Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA* 2006; 295: 1549–1555.
7. Centers for Disease Control and Prevention. 2007; National Diabetes Fact Sheet. Available at: <http://apps.nccd.cdc.gov/ddtstrs/factsheet.aspx>. Accessed September 2008.
8. Santry HP, Gillen DL, Lauderdale DS. Trends in bariatric surgical procedures. *JAMA* 2005; 294: 1909–1917.
9. Herron DM. The surgical management of severe obesity. *Mt Sinai J Med* 2004; 71: 63–71.
10. Parikh MS, Fielding GA, Ren CJ. U.S. experience with 749 laparoscopic adjustable gastric bands: intermediate outcomes. *Surg Endosc* 2005; 19: 1631–1635.
11. Cummings DE, Weigle DS, Frayo RS, et al. Plasma ghrelin levels after diet-induced weight loss or gastric bypass surgery. *N Engl J Med* 2002; 346: 1623–1630.
12. Schauer PR, Burguera B, Ikramuddin S, et al. Effect of laparoscopic Roux-en Y gastric bypass on T2DM mellitus. *Ann Surg* 2003; 238: 467–484.
13. Sjostrom L, Lindroos A-K, Peltonen M, et al. Lifestyle, diabetes and cardiovascular risk factors 10 years after bariatric surgery. *N Engl J Med* 2004; 351: 2683–2693.
14. Sjöström CD, Peltonen M, Wedel H, et al. Differentiated long-term effects of intentional weight loss on diabetes and hypertension. *Hypertension* 2000; 36: 20–25.
15. Sjöström CD, Lissner L, Wedel H, et al. Reduction in incidence of diabetes, hypertension and lipid disturbances after intentional weight loss induced by bariatric surgery: the SOS Intervention Study. *Obes Res* 1999; 7: 477–484.
16. Torquati A, Luthi R, Abumrad N, Richards WO. Is Roux-en-Y gastric bypass surgery the most effective treatment for T2DM mellitus in morbidly obese patients? *J Gastrointest Surg* 2005; 9: 1112–1116.
17. MacDonald KG Jr, Long SD, Swanson MS, et al. The gastric bypass operation reduces the progression and mortality of non-insulin-dependent diabetes mellitus. *J Gastrointest Surg* 1997; 1: 213–220.
18. Sugerma HJ, Wolfe LG, Sica DA, Clore JN. Diabetes and hypertension in severe obesity and effects of gastric bypass-induced weight loss. *Ann Surg* 2003; 237: 751–756.
19. Pories WJ, MacDonald KG Jr, Flickinger EG, et al. Is type II diabetes mellitus (NIDDM) a surgical disease? *Ann Surg* 1992; 215: 633–642.
20. Long SD, O'Brien K, MacDonald KG Jr, et al. Weight loss in severely obese subjects prevents the progression of impaired glucose tolerance to type II

- diabetes. A longitudinal interventional study. *Diabetes Care* 1994; 17: 372–375.
21. Marceau P, Biron S, Hould FS, et al. Duodenal switch: long-term results. *Obes Surg* 2007; 17: 1421–1430.
 22. Marinari GM, Papadia FS, Briatore L, et al. T2DM and weight loss following biliopancreatic diversion for obesity. *Obes Surg* 2006; 16: 1440–1444.
 23. Scopinaro N, Marinari GM, Camerini GB, et al. Specific effects of biliopancreatic diversion on the major components of metabolic syndrome: a long-term follow-up study. *Diabetes Care* 2005; 10: 2406–2411.
 24. Buchwald H, Avidor Y, Braunwald E, et al. Bariatric surgery: a systematic review and meta-analysis. *JAMA* 2004; 292: 1724–1737.
 25. Polyzogopoulou EV, Kalfarentzos F, Vagenakis AG, et al. Restoration of euglycemia and normal acute insulin response to glucose in obese subjects with T2DM following bariatric surgery. *Diabetes* 2003; 52: 1098–1103.
 26. Scopinaro N, Papadia F, Camerini G, et al. A comparison of a personal series of biliopancreatic diversion and literature data on gastric bypass help to explain the mechanisms of resolution of T2DM by the two operations. *Obes Surg* 2008; 18: 1035–1038.
 27. Dixon JB, O'Brien PE. Health outcomes of severely obese type 2 diabetic subjects 1 year after laparoscopic adjustable gastric banding. *Diabetes Care* 2002; 25: 358–363.
 28. Hess DS, Hess DW. Biliopancreatic diversion with a duodenal switch. *Obes Surg* 1998; 8: 267–282.
 29. Silecchia G, Boru C, Pecchia A, et al. Effectiveness of laparoscopic sleeve gastrectomy (first stage of biliopancreatic diversion with duodenal switch) on co-morbidities in super-obese high-risk patients. *Obes Surg* 2006; 16: 1138–1144.
 30. Gan SS, Talbot ML, Jorgensen JO. Efficacy of surgery in the management of obesity-related T2DM mellitus. *ANZ J Surg* 2007; 77: 958–962.
 31. Vidal J, Ibarzabal A, Romero F, et al. T2DM mellitus and the metabolic syndrome following sleeve gastrectomy in severely obese subjects. *Obes Surg* 2008; 18: 1077–1082.
 32. Balsiger BM, Poggio JL, Mai J, et al. Ten and more years after vertical banded gastroplasty as primary operation for morbid obesity. *Gastrointest Surg* 2000; 4: 598–605.
 33. Dixon JB, O'Brien PE, Playfair J, et al. Adjustable gastric banding and conventional therapy for T2DM: a randomized controlled trial. *JAMA* 2008; 299: 316–323.
 34. Ponce J, Haynes B, Paynter S, et al. Effect of lap-band-induced weight loss on T2DM mellitus and hypertension. *Obes Surg* 2004; 14: 1335–1342.
 35. Wickremesekera K, Miller G, Naotunne TD, et al. Loss of insulin resistance after Roux-en-Y gastric bypass surgery: a time course study. *Obes Surg* 2005; 15: 474–481.
 36. Scopinaro N, Adami GF, Marinari GM, et al. Biliopancreatic diversion. *World J Surg* 1998; 22: 936–946.
 37. Hickey MS, Pories WJ, MacDonald KG Jr, et al. A new paradigm for T2DM mellitus: could it be a disease of the foregut? *Ann Surg* 1998; 227: 637–643.
 38. Rubino F, Forgione A, Cummings DE, et al. The mechanism of diabetes control after gastrointestinal bypass surgery reveals a role of the proximal small intestine in the pathophysiology of T2DM. *Ann Surg* 2006; 244: 741–749.
 39. Lima J, Helena L, Oliveira S, et al. Rapid resolution of diabetes after gastric bypass. *Obes Surg* 2005; 15: 448–449.
 40. Kellum JM, Kuemmerle JF, O'Dorisio TM, et al. Gastrointestinal hormone responses to meals before and after gastric bypass and vertical banded gastroplasty. *Ann Surg* 1990; 6: 763–770.
 41. Greco AV, Mingrone G, Giancaterini A, et al. Insulin resistance in morbid obesity: reversal with intramyocellular fat depletion. *Diabetes* 2002; 51: 144–1451.
 42. Brodin RE, LaMarca LB, Kenler HA, et al. Malabsorptive gastric bypass in patients with superobesity. *J Gastrointest Surg* 2002; 6: 195–203.
 43. Faraj M, Jones P, Sniderman AD, et al. Enhanced dietary fat clearance in postobese women. *J Lipid Res* 2001; 42: 571–580.
 44. MacLean LD, Rhode BM, Nohr CW. Long- or short-limb gastric bypass? *J Gastrointest Surg* 2001; 5: 525–530.
 45. Sugerman HJ, Londrey GL, Kellum JM, et al. Weight loss with vertical banded gastroplasty and Roux-Y gastric bypass for morbid obesity with selective versus random assignment. *Am J Surg* 1989; 157: 93–102.
 46. Korner J, Bessler M, Cirilo LJ, et al. Effects of Roux-en-Y gastric bypass surgery on fasting and postprandial concentrations of plasma ghrelin, peptide YY, and insulin. *J Clin Endocrinol Metab* 2005; 90: 359–365.
 47. Kreymann B, Williams G, Ghatei MA, et al. Glucagon-like peptide-1 7–36: a physiological incretin in man. *Lancet* 1987; 2: 1300–1304.
 48. Orskov C. Glucagon-like peptide-1, a new hormone of the entero-insular axis. *Diabetologia* 1992; 35: 701–711.
 49. Drucker DJ. Gut adaptation and the glucagon-like peptides. *Gut* 2002; 50: 428–435.
 50. Flint A, Raben A, Ersbøll AK, et al. The effect of physiological levels of glucagon-like peptide-1 on appetite, gastric emptying, energy and substrate metabolism in obesity. *Int J Obes Relat Metab Disord* 2001; 25: 781–792.
 51. Willms B, Werner J, Holst JJ, et al. Gastric emptying, glucose responses, and insulin secretion after a liquid test meal: effects of exogenous glucagon-like peptide-1 (GLP-1)-(7–36) amide in type 2 (noninsulin-dependent) diabetic patients. *J Clin Endocrinol Metab* 1996; 81: 327–332.
 52. Gutniak M, Orskov C, Holst JJ, et al. Antidiabetogenic effect of glucagon-like peptide-1 (7–36) amide in normal subjects and patients with diabetes mellitus. *N Engl J Med* 1992; 326: 1316–1322.
 53. Mason EE. Ileal [correction of ilial] transposition and enteroglucagon/GLP-1 in obesity (and diabetic?) surgery. *Obes Surg* 1999; 9: 223–228.
 54. Brubaker PL, Drucker DJ. Minireview: glucagon-like peptides regulate cell proliferation and apoptosis in the pancreas, gut, and central nervous system. *Endocrinology* 2004; 145: 2653–2659.
 55. Service FJ, Thompson GB, Service FJ, et al. Hyperinsulinemic hypoglycemia with nesidioblastosis after gastric-bypass surgery. *N Engl J Med* 2005; 353: 249–254.

56. Vilsbøll T, Krarup T, Deacon CF, et al. Reduced postprandial concentrations of intact biologically active glucagon-like peptide 1 in type 2 diabetic patients. *Diabetes* 2001; 50: 609–613.
57. Nauck M, Stöckmann F, Ebert R, et al. Reduced incretin effect in type 2 (non-insulin-dependent) diabetes. *Diabetologia* 1986; 29: 46–52.
58. Toft-Nielsen MB, Madsbad S, Holst JJ. Determinants of the effectiveness of glucagon-like peptide-1 in T2DM. *J Clin Endocrinol Metab* 2001; 86: 3853–3860.
59. Muscelli E, Mari A, Casolaro A, et al. Separate impact of obesity and glucose tolerance on the incretin effect in normal subjects and type 2 diabetic patients. *Diabetes* 2008; 57: 1340–1348.
60. Freyse EJ, Becher T, El-Hag O, et al. Blood glucose lowering and glucagonostatic effects of glucagon-like peptide I in insulin-deprived diabetic dogs. *Diabetes* 1997; 46: 824–828.
61. Zander M, Madsbad S, Madsen JL, et al. Effect of 6-week course of glucagon-like peptide 1 on glycaemic control, insulin sensitivity, and beta-cell function in T2DM: a parallel-group study. *Lancet* 2002; 359: 824–830.
62. Sarson DL, Scopinaro N, Bloom SR. Gut hormone changes after jejunoileal (JIB) or biliopancreatic (BPB) bypass surgery for morbid obesity. *Int J Obes* 1981; 5: 471–480.
63. Kellum JM, Kuemmerle JF, O'Dorisio TM, et al. Gastrointestinal hormone responses to meals before and after gastric bypass and vertical banded gastroplasty. *Ann Surg* 1990; 211: 763–770.
64. Näslund E, Backman L, Holst JJ, et al. Importance of small bowel peptides for the improved glucose metabolism 20 years after jejunoileal bypass for obesity. *Obes Surg* 1998; 8: 253–260.
65. Valverde I, Puente J, Martín-Duce A, et al. Changes in glucagon-like peptide-1 (GLP-1) secretion after biliopancreatic diversion or vertical banded gastroplasty in obese subjects. *Obes Surg* 2005; 15: 387–397.
66. Clements RH, Gonzalez QH, Long CI, et al. Hormonal changes after Roux-en Y gastric bypass for morbid obesity and the control of type-II diabetes mellitus. *Am Surg* 2004; 70: 1–4.
67. Le Roux CW, Aylwin SJ, Batterham RL, et al. Gut hormone profiles following bariatric surgery favor an anorectic state, facilitate weight loss, and improve metabolic parameters. *Ann Surg* 2006; 243: 108–114.
68. Le Roux CW, Welbourn R, Werling M, et al. Gut hormones as mediators of appetite and weight loss after Roux-en-Y gastric bypass. *Ann Surg* 2007; 246: 780–785.
69. Laferrère B, Heshka S, Wang K, et al. Incretin levels and effect are markedly enhanced 1 month after Roux-en-Y gastric bypass surgery in obese patients with T2DM. *Diabetes Care* 2007; 30: 1709–1716.
70. Morínigo R, Moizé V, Musri M, et al. Glucagon-like peptide-1, peptide YY, hunger, and satiety after gastric bypass surgery in morbidly obese subjects. *J Clin Endocrinol Metab* 2006; 91: 1735–1740.
71. Morínigo R, Lacy AM, Casamitjana R, et al. Changes in glucose tolerance following gastric bypass surgery in morbidly obese subjects. *Obes Surg* 2006; 16: 1594–15601.
72. Bloom SR, Polak JM. Gut hormones. *Adv Clin Chem* 1980; 21: 177–244.
73. Adrian TE, Ferri GL, Bacarese-Hamilton AJ, et al. Human distribution and release of a putative new gut hormone, peptide YY. *Gastroenterology* 1985; 89: 1070–1077.
74. Neary NM, Small CJ, Druce MR, et al. Peptide YY3–36 and glucagon-like peptide-17–36 inhibit food intake additively. *Endocrinology* 2005; 146: 5120–5127.
75. Nauck MA, Heimesaat MM, Orskov C, et al. Preserved incretin activity of glucagon-like peptide 1 [7–36 amide] but not of synthetic human gastric inhibitory polypeptide in patients with type-2 diabetes mellitus. *J Clin Invest* 1993; 91: 301–307.
76. Elahi D, McAloon-Dyke M, Fukagawa NK, et al. The insulinotropic actions of glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (7–37) in normal and diabetic subjects. *Regul Pept* 1994; 51: 63–74.
77. Jorde R, Burhol PG, Johnson JA. The effect of jejunioleal bypass on postprandial release of plasma gastric inhibitory polypeptide (GIP). *Scand J Gastroenterol* 1981; 16: 313–319.
78. Lauritsen KB, Christensen KC, Stokholm KH. Gastric inhibitory polypeptide (GIP) release and incretin effect after oral glucose in obesity and after jejunioleal bypass. *Scand J Gastroenterol* 1980; 15: 489–495.
79. Sirinek KR, O'Dorisio TM, Hill D, et al. Hyperinsulinism, glucose-dependent insulinotropic polypeptide, and the enteroinsular axis in morbidly obese patients before and after gastric bypass. *Surgery* 1986; 100: 781–787.
80. Rubino F, Marescaux J. Effect of duodenal-jejunal exclusion in a non-obese animal model of T2DM: a new perspective for an old disease. *Ann Surg* 2004; 239: 1–11.
81. Rodieux F, Giusti V, D'Alessio DA, et al. Effects of gastric bypass and gastric banding on glucose kinetics and gut hormone release. *Obesity* 2008; 16: 298–305.
82. Batterham RL, Cohen MA, Ellis SM, et al. Inhibition of food intake in obese subjects by peptide YY3–36. *N Engl J Med* 2003; 349: 941–948.
83. Batterham RL, Cowley MA, Small CJ, et al. Gut hormone PYY(3–36) physiologically inhibits food intake. *Nature* 2002; 418: 650–654.
84. Näslund E, Grybäck P, Hellström PM, et al. Gastrointestinal hormones and gastric emptying 20 years after jejunioleal bypass for massive obesity. *Int J Obes Relat Metab Disord* 1997; 21: 387–392.
85. Strader AD, Vahl TP, Jandacek RJ, et al. Weight loss through ileal transposition is accompanied by increased ileal hormone secretion and synthesis in rats. *Am J Physiol Endocrinol Metab* 2005; 288: E447–E453.
86. Szczówka J, Tatemoto K, Rajamäki G, et al. Effects of PYY and PP on endocrine pancreas. *Acta Physiol Scand* 1983; 119: 123–126.
87. Faraj M, Havel PJ, Phélis S, et al. Plasma acylation-stimulating protein, adiponectin, leptin, and ghrelin before and after weight loss induced by gastric bypass surgery in morbidly obese subjects. *J Clin Endocrinol Metab* 2003; 88: 1594–1602.
88. Emilsson V, Liu YL, Cawthorne MA, et al. Expression of the functional leptin receptor mRNA in pancreatic islets and direct inhibitory action of leptin on insulin secretion. *Diabetes* 1997; 46: 313–316.
89. Kieffer TJ, Heller RS, Leech CA, et al. Leptin suppression of insulin secretion by the activation of

- ATP-sensitive K⁺ channels in pancreatic beta-cells. *Diabetes* 1997; 46: 1087–1093.
90. Seufert J, Kieffer TJ, Leech CA, et al. Leptin suppression of insulin secretion and gene expression in human pancreatic islets: implications for the development of adipogenic diabetes mellitus. *J Clin Endocrinol Metab* 1999; 84: 670–676.
 91. Lam NT, Cheung AT, Riedel MJ, et al. Leptin reduces glucose transport and cellular ATP levels in INS-1 beta-cells. *J Mol Endocrinol* 2004; 32: 415–424.
 92. Covey SD, Wideman RD, McDonald C, et al. The pancreatic β cell is a key site for mediating the effects of leptin on glucose homeostasis. *Cell Metab* 2006; 4: 291–302.
 93. Wauters M, Considine RV, Yudkin JS, et al. Leptin levels in T2DM: associations with measures of insulin resistance and insulin secretion. *Horm Metab Res* 2003; 35: 92–96.
 94. Kerner J, Inabnet W, Conwell IM, et al. Differential effects of gastric bypass and banding on circulating gut hormone and leptin levels. *Obesity* 2006; 14: 1553–1561.
 95. Infanger D, Baldinger R, Branson R, et al. Effect of significant intermediate-term weight loss on serum leptin levels and body composition in severely obese subjects. *Obes Surg* 2003; 13: 879–888.
 96. Holdstock C, Engström BE, Ohrvall M, et al. Ghrelin and adipose tissue regulatory peptides: effect of gastric bypass surgery in obese humans. *J Clin Endocrinol Metab* 2003; 88: 3177–3183.
 97. Molina A, Vendrell J, Gutiérrez C, et al. Insulin resistance, leptin and TNF-alpha system in morbidly obese women after gastric bypass. *Obes Surg* 2003; 13: 615–621.
 98. Yamauchi T, Kamon J, Minokoshi Y, et al. Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. 1. *Nat Med* 2002; 8: 1288–1295.
 99. Matsuzawa Y, Funahashi T, Nakamura T. Molecular mechanism of metabolic syndrome X: contribution of adipocytokines adipocyte-derived bioactive substances. *Ann N Y Acad Sci* 1999; 892: 146–154.
 100. Ballantyne GH, Gumbs A, Modlin IM. Changes in insulin resistance following bariatric surgery and the adipoinular axis: role of the adipocytokines, leptin, adiponectin and resistin. *Obes Surg* 2005; 15: 692–699.
 101. Yang WS, Lee WJ, Funahashi T, et al. Plasma adiponectin levels in overweight and obese Asians. *Obes Res* 2002; 10: 1104–1110.
 102. Kubota N, Terauchi Y, Yamauchi T, et al. Disruption of adiponectin causes insulin resistance and neo-intimal formation. *J Biol Chem* 2002; 277: 25863–25866.
 103. Spranger J, Kroke A, Möhlig M, et al. Adiponectin and protection against T2DM mellitus. *Lancet* 2003; 361: 226–228.
 104. Hanusch-Enserer U, Brabant G, Roden M. Ghrelin concentrations in morbidly obese patients after adjustable gastric banding. *N Engl J Med* 2003; 348: 2159–2160.
 105. Pories WJ, Albrecht RJ. Etiology of type II diabetes mellitus: role of the foregut. *World J Surg* 2001; 25: 527–531.
 106. Rubino F, Gagner M. Potential of surgery for curing T2DM mellitus. *Ann Surg* 2002; 236: 554–559.
 107. Wang TT, Hu SY, Gao HD, et al. Ileal transposition controls diabetes as well as modified duodenal jejunal bypass with better lipid lowering in a nonobese rat model of type II diabetes by increasing GLP-1. *Ann Surg* 2008; 247: 968–975.
 108. Pacheco D, de Luis DA, Romero A, et al. The effects of duodenal-jejunal exclusion on hormonal regulation of glucose metabolism in Goto-Kakizaki rats. *Am J Surg* 2007; 194: 221–224.
 109. Mason EE. The mechanisms of surgical treatment of T2DM. *Obes Surg* 2005; 15: 459–461.
 110. Patrity A, Facchiano E, Sanna A, et al. The enteroinsular axis and the recovery from T2DM after bariatric surgery. *Obes Surg* 2004; 14: 840–848.
 111. Cummings DE, Overduin J, Foster-Schubert KE. Gastric bypass for obesity: mechanisms of weight loss and diabetes resolution. *J Clin Endocrinol Metab* 2004; 89: 2608–2615.
 112. Patrity A, Aisa MC, Annetti C, et al. How the hindgut can cure T2DM. Ileal transposition improves glucose metabolism and beta-cell function in Goto-Kakizaki rats through an enhanced proglucagon gene expression and L-cell number. *Surgery* 2007; 142: 74–85.
 113. Dolan K, Bryant R, Fielding G. Treating diabetes in the morbidly obese by laparoscopic gastric banding. *Obes Surg* 2003; 13: 439–443.
 114. O'Brien PE, Dixon JB. Lap-band: outcomes and results. *J Laparoendosc Adv Surg Tech A* 2003; 13: 265–270.
 115. Cottam D, Qureshi FG, Mattar SG, et al. Laparoscopic sleeve gastrectomy as an initial weight-loss procedure for high-risk patients with morbid obesity. *Surg Endosc* 2006; 20: 859–863.
 116. Vidal J, Ibarzabal A, Nicolau J, et al. Short-term effects of sleeve gastrectomy on T2DM mellitus in severely obese subjects. *Obes Surg* 2007; 17: 1069–1074.
 117. Scopinaro N, Marinari GM, Camerini GB, et al. Specific effects of biliopancreatic diversion on the major components of metabolic syndrome: a long-term follow-up study. *Diabetes Care* 2005; 28: 2406–2411.
 118. White S, Brooks E, Jurikova L, et al. Long-term outcomes after gastric bypass. *Obes Surg* 2005; 15: 155–163.
 119. Pontiroli AE, Pizzocri P, Librenti MC, et al. Laparoscopic adjustable gastric banding for the treatment of morbid (grade 3) obesity and its metabolic complications: a three-year study. *J Clin Endocrinol Metab* 2002; 87: 3555–3561.
 120. Sjöström L, Narbro K, Sjöström CD, et al. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med* 2007; 357: 741–752.
 121. Pontiroli AE, Folli F, Paganelli M, et al. Laparoscopic gastric banding prevents T2DM and arterial hypertension and induces their remission in morbid obesity: a 4-year case-controlled study. *Diabetes Care* 2005; 28: 2703–2709.
 122. O'Brien PE, Dixon JB, Laurie C, et al. Treatment of mild to moderate obesity with laparoscopic adjustable gastric banding or an intensive medical program: a randomized trial. *Ann Intern Med* 2006; 144: 625–633.
 123. Van Dielen FM, Soeters PB, de Brauw LM, et al. Laparoscopic adjustable gastric banding versus open vertical banded gastroplasty: a prospective randomized trial. *Obes Surg* 2005; 15: 1292–1298.

EVIDENCE-BASED MEDICINE

Just a Spoonful of Sugar: Does Presenting Hyperglycemia Affect Prognosis in Nondiabetics with Acute Myocardial Infarction

Laurie Edelman, Thomas McGinn, and Deborah Korenstein

Mount Sinai School of Medicine
New York, NY

This is the fifth in a series of articles in the *Mount Sinai Journal of Medicine* on evidence-based medicine. This series will emphasize applications of evidence-based medicine concepts and procedures to clinical practice. The articles will focus on the retrieval, interpretation, and application of evidence related to a current clinical problem. The discussion will be based on the following case.

THE CASE

A 62-year-old woman presents to the emergency department with an acute ST-elevation myocardial infarction. She has hypertension and is a former smoker but denies any history of diabetes. She is found to have a random glucose level of 208 mg/dL on her admission chemistry panel. She is seen by cardiology and undergoes emergent cardiac catheterization with a successful intervention.

She recovers and returns for follow-up. She has been taking no diabetic medication, and her fasting blood glucose is 89 mg/dL with a hemoglobin A1c of 5.6%. After this traumatic experience, she is concerned about her mortality. She asks her physician

if the elevated glucose in the emergency department confers a worse prognosis over the longer term.

FOCUSED CLINICAL QUESTION

To evaluate whether hyperglycemia in the emergency department affects this patient's prognosis, the physician decides to review the medical literature. She begins by formulating a focused clinical question consisting of 4 elements: the patient (P), the intervention (I), the comparison (C), and the outcome (O). This is called a PICO question.¹ The explicit inclusion of these 4 components allows for a more efficient and directed search of the literature and helps to clarify the exact problem of interest. In this case, the question's components are as follows: In patients with acute myocardial infarction (P), does hyperglycemia at presentation (I), in comparison with normal glucose (C), predict an increased mortality rate (O)? The PICO format was designed to structure questions about treatment, but it works well for this prognosis question as well. In future installments in this series, we will discuss questions about diagnosis, which are more difficult to place in the PICO format.

Address Correspondence to:

Laurie Edelman

Mount Sinai School of Medicine
New York, NY
Email: laurie.edelman@
mountsinai.org

SEARCH OF THE LITERATURE

The physician goes to PubMed and begins searching for answers to her clinical question by using the prognosis filter in Clinical Queries. Clinical Queries in PubMed uses filters (extra terms in the search) to find articles in a specific category (eg, therapy, diagnosis,

prognosis, and systematic review). It can be a useful tool for finding articles, although the prognosis filter is less effective than some of the other filters. She enters the search terms "hyperglycemia" and "acute MI." This search reveals 15 articles. She selects the eighth article, which is entitled "The Impact of Admission Plasma Glucose on Long-Term Mortality After STEMI and NSTEMI Myocardial Infarction."² The study addressed the topic but had a small sample size and was limited to a Veterans Affairs population, so the physician clicks Related Articles. Related Articles provides a list of papers on similar topics and can be helpful when the original search does not yield the best articles. She then selects the article entitled "Is Blood Glucose an Independent Predictor of Mortality in Acute Myocardial Infarction in the Thrombolytic Era,"³ which seems appropriate and adequately large, so she decides to read that article to answer her question.

Validity Criteria for Prognosis

The study used a large population-based database from Nova Scotia. Within the database, investigators identified 1664 consecutive patients hospitalized with acute myocardial infarction. The patients were then stratified on the basis of a previous diagnosis of diabetes and their admission blood glucose, which was classified as less than 198 mg/dL or greater than 198 mg/dL. These patients were followed for 12 months with an endpoint of mortality. The study set out to determine whether elevated admission glucose was associated with higher mortality in diabetics and nondiabetics.

In terms of the major validity issues,⁴ the following 4 points should be considered.

First, the sample of patients is **representative** of patients hospitalized with acute myocardial infarction. To accurately determine the prognosis of a given population of patients, a study needs to capture patients who are representative of that population. If the population under study is unique in some fundamental way, then bias may be introduced, and the study results may not be reliable. When evaluating a prognosis study, one should think about whether the process of selecting the population might have systematically favored a certain subgroup of patients. For example, in a study of the prognosis of arthritis, if the patients were recruited from a rheumatology practice, they were likely to have severe disease and may not be representative of the entire population of patients with arthritis. This study involved all patients hospitalized with acute myocardial infarction in Nova Scotia during a 1-year period, so we can expect that they are representative of myocardial infarction

patients, at least those who are hospitalized in Nova Scotia.

Second, the patients were sufficiently **homogeneous with respect to prognostic risk**. This issue is of central importance in questions of prognosis. The patients enrolled in a prognosis trial, known as the inception cohort, need to be sufficiently homogeneous that it makes sense to study them as a group. If the patients in the inception cohort are wildly variable in terms of their prognostic risk, it will be difficult and potentially misleading to draw conclusions about them as a group. For example, in a study of severe pain, it would not make sense to group patients with acute traumatic fracture with patients with diabetic neuropathy, although both have pain. In this case, the inception cohort consisted of consecutive patients with acute myocardial infarction, and this was an adequately homogeneous group for the purposes of the study.

Even in the most homogeneous group of patients, individuals will differ with respect to various risk factors other than the specific ones under investigation. Investigators must consider these other factors that affect prognosis in their analysis. In this study, the primary question related the impact of blood sugar on prognosis, but the patients with high glucose differed from those with normal glucose in terms of many prognostic factors, such as age and comorbidities, so the investigators had to demonstrate whether glucose was actually the important prognostic factor. There are several statistical methods for doing this, but the simplest is an adjusted analysis, in which investigators determine the impact of blood glucose separately in, for instance, older and younger patients. In this way, they can separate the effects of the various variables. In this study, the investigators performed a multivariate analysis, looking separately at a number of important variables.

Third, **follow-up** was sufficiently long and complete. This concept contains 2 issues: whether follow-up of the population is sufficiently long to find the outcome of interest and whether it is complete, accounting for all or most of the enrolled patients. The appropriate length of follow-up will depend on the issue at hand. In this case, the duration of follow-up was 12 months, which seems long enough because acute myocardial infarction is associated with high 1-year mortality. In this study, all patients received complete follow-up, likely because of the way in which the original data were collected.

Fourth, the **outcome criteria** were objective and unbiased. It is important to consider the nature of the outcome criteria when one is evaluating a

prognosis article in terms of both its clinical significance and its objectivity. This study used death, as defined by the Nova Scotia Vital Statistics Registry, as the endpoint. This province-wide registry is constantly updated and has multiple internal controls. The study team further verified death through a review of obituaries and continuous monitoring of patient medical records. Death is an unbiased criterion, and the investigators appeared to have reliably determined which patients died.

RESULTS

The study examined known diabetics as well as nondiabetics, but because the physician's current patient was not known to be diabetic, the physician chose to look only at the nondiabetic patients. The nondiabetic patients were divided into 2 groups on the basis of a presenting random glucose level greater or less than 198. Not surprisingly, there were other differences between the groups, including creatinine, age, history of heart failure, hyperlipidemia, smoking, and medications at baseline. In terms of the hospital course in the 2 groups, more hyperglycemic patients had heart failure (32.6% versus 13.3%), but interventions in the groups were similar. The investigators performed a multivariate analysis of predictors of in-hospital mortality; this means that they looked at a series of factors to determine the relative contribution of each to the outcome. Hyperglycemia at admission was associated with an odds ratio of 2.44 for in-hospital mortality; this means that a hyperglycemic patient dying in the hospital were 2.44 times the odds for a euglycemic patient. Not surprisingly, there were other factors that predicted in-hospital mortality as well, including peripheral vascular disease, age, and previous myocardial infarction. The authors also presented 1-year mortality rates in the 2 groups, showing that 8.7% of hyperglycemic nondiabetics died in 1 year versus 3.1% of nonhyperglycemics. However, there were no adjustments made for other prognostic factors, and because those prognostic factors differed substantially between the groups, it is difficult to interpret the 1-year mortality difference, which may well be due to factors other than hyperglycemia.

APPLICABILITY

The physician decides that the study applies to her patient and that it has fairly good validity, but she has some difficulty applying the results. She feels that the

in-hospital mortality data are reliable because they take other prognostic factors into account, but her patient has already survived her hospitalization. The higher 1-year mortality is more difficult to interpret because it may be due to differences in other factors. She decides to tell her patient that she is at risk for increased in-hospital mortality but that it is not clear what her risk is in the long term. She decides, however, to screen the patient for diabetes regularly, manage her other risk factors aggressively, and follow her closely for the future development of diabetes.

CONCLUSIONS

The case illustrates several important points:

- Try PubMed's Clinical Queries when looking for articles about prognosis, and click on Related Articles when the reference that is found is close but not perfect.
- It is important to assess the validity of articles about prognosis before interpreting the results. Focus on issues of the inception cohort, the length of follow-up, and the outcome criteria used.
- In prognosis articles, it is important for investigators to control for other prognostic factors that may differ between groups. Look for assurances in the article that this was done and look for adjusted odds ratios or relative risks as the most accurate reflection of the contribution of a particular risk factor.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12-A13.
2. Dirkali A, van der Ploeg T, Nangrahary M, et al. The impact of admission plasma glucose on long-term mortality after STEMI and NSTEMI myocardial infarction. *Int J Cardiol* 2007; 121: 215-217.
3. Wahab NN, Cowden EA, Pearce NJ, et al. Is blood glucose an independent predictor of mortality in acute myocardial infarction in the thrombolytic era? *J Am Coll Cardiol* 2002; 40: 1748-1754.
4. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1993; 270: 2598-2601.

ETHICS

Dilemmas in Surgery: Medical Ethics Education in Surgery Rotation

Nada Gligorov, PhD, Pippa Newell, MD, Jason Altilio, MPhil, Mike Collins, MPhil,
Amanda Favia, Leah Rosenberg, and Rosamond Rhodes, PhD

Mount Sinai School of Medicine, New York, NY

Mount Sinai School of Medicine provides its students with a 4-year curriculum in medical ethics that is horizontally and vertically integrated into the educational program. In the first 2 years, as part of the Arts and Science of Medicine (ASM) course (ASM 1 and ASM 2), students participate in 6 ethics modules, the first 5 of which are designed to provide them with an understanding of the basic concepts of medical ethics. Sessions focus on professionalism, fiduciary responsibility, truth telling and informed consent, autonomy, surrogate decision making, justice, and research ethics. Through readings, minilectures, panels, sample cases, and small group discussions, the sessions provide students with the vocabulary and understanding to enable them to navigate the ethical issues of clinical practice. The final module of the second year presents students with a model of clinical moral reasoning to be used in the resolution of ethical dilemmas.¹

In addition to the medical ethics modules of the first 2 years, medical ethics education is continued in the clinical years with ethics sessions that are incorporated into each clinical clerkship, and students' competence in medical ethics is assessed as part of each of their 2 comprehensive assessments. The clinical clerkships involve clinical experience at Mount Sinai and our many affiliated hospitals

and also some time for didactic sessions, including medical ethics. Clerkship sessions are cotaught in small groups by a member of the clinical department and a member of the ethics faculty in order to bring together the relevant expertise. Most sessions are built around a single topic (eg, giving bad news or refusal of treatment), with readings provided and sample cases prepared for leading the discussions. These sessions aim at improving the students' understanding of key concepts, promoting their ability to employ the concepts in a variety of settings, and developing the habit of discussing ethical issues with peers.

ETHICS PROJECT

In July 2001, we revised the ethics session for the surgery clerkship by instituting a new activity, an ethics project. In this activity, instead of addressing a particular topic, we chose to focus on fostering skills that are essential for competence in medical ethics. Key skills that we want our students to develop include the following:

- The ability to identify an ethical issue.
- The ability to use the bioethics literature in addressing the issue.
- The ability to lead a discussion of the matter with peers.
- The ability to resolve a dilemma by using the model of clinical moral reasoning.
- The ability to offer reasons for a position that others can appreciate and accept.

With those goals in mind, we ask the students in the surgery clerkship to complete an ethics project. At the beginning of the surgery clerkship, students

Address Correspondence to:

Nada Gligorov

Mount Sinai School of Medicine
New York, NY

Email: nada.gligorov@exchange.mssm.edu

Table 1. *Central Concepts in Students' Ethics Project Presentations in the Surgery Clerkship.*

Central Ethical Concept	Number of Projects
Informed consent (including scope)	18
Truth telling (medical errors; learning on patients)	39
Capacity/surrogacy issues	21
Professionalism (time management; staff interaction)	17
Beneficence	13
Duty to care	7
Autonomy	7
Justice	7
Paternalism	5
Nonjudgmental regard	3
Other	5
Total responses	172

receive a short guide to the bioethics literature and other pertinent electronic resources. The clerkship handbook instructs them to be attentive, from day 1 of the clerkship onward, to clinical situations raising ethical issues. They are instructed to search the bioethics literature on their topic and present a case of their choice for their small group ethics session. The presentation is to include a summary of the case, a presentation of relevant literature, and questions for leading the group in a discussion of the case. The students then meet in small groups with a surgeon and an ethicist.

The topics addressed in the last 2 years of ethics project case presentations have been recorded by the ethics faculty members who cofacilitate the small group discussions. Although the cases cover a wide variety of subjects, from duty to provide care to just allocation of resources and medical futility, 3 dominant topics have emerged. The students most frequently present cases addressing issues related to (1) the scope of informed consent, (2) truth telling with respect to the occurrence and disclosure of medical errors or the role of learners, and (3) decisional capacity (Table 1).

CASES FROM ETHICS PROJECT PRESENTATIONS

To more fully explain the scope of the ethics project in the surgery clerkship and to illustrate the kinds of issues that our students have chosen to present, we offer 3 cases exemplifying some of the most common topics discussed in student projects. Each case is followed with an ethical analysis that

captures key points that arose in the student case discussion.

Scope of Informed Consent Case

A 47-year-old male was admitted to the hospital for the removal of a blockage in his colon that the physicians believed to be caused by chronic diverticulitis. Prior to the surgery, the patient gave consent to laparoscopic surgery to remove the blockage. The patient was also informed that the procedure would be converted to open surgery if it became necessary. During the procedure, the surgeons decided to convert to open surgery and incidentally discovered a tumor in the appendix that was unrelated to the blockage. The surgeons discussed removing the tumor and decided to proceed with its removal without obtaining additional consent.

Discussion

The question in this case is whether the surgeon should try to help the patient by removing the tumor and avoiding a second surgery or show respect for his autonomy by first obtaining his informed consent to the extra procedure. Performing the resection without the patient's authorization would be paternalistic. The question then is whether the paternalistic intervention would be justified.

One relevant concept is the scope of the informed consent that was obtained. The patient signed a consent form prior to the surgery, but, of course, it did not specifically mention the removal of an unexpected tumor. Should the doctors remove the tumor, assuming that it would be medically necessary in the long run, or should they wait to speak to the patient and ascertain his wishes?

What is needed is a determination of the risks and benefits associated with leaving the tumor or immediately removing it. If the risks of leaving the tumor are extremely high, then the removal of the tumor is justified. For example, if the removal of the tumor is emergent or potentially life-saving, the surgeons are justified in assuming that most people who had agreed to surgery to remove a colon blockage would also agree to the removal of a life-threatening tumor. Also, if an additional surgery would be too risky for this patient, then the removal could be justified if we assume that most people would not wish to incur further significant and unnecessary risks. In all these cases, the justification for removing the tumor does not rest on explicit consent from the patient, but on the assumption that

most reasonable people faced with similar associated risks and benefits would agree that the removal of the tumor is justified. Another way of considering how to proceed would involve imagining how a person awaking from anesthesia would react to the fact that a tumor (about which he knew nothing) had been removed without his consent. If it can be assumed that the patient would be happy with the decision, because most people in this circumstance would be, then we could invoke the concept of future-oriented consent to justify proceeding with the additional procedure.² If an authorized surrogate (ie, a healthcare Proxy) is available for a discussion during the surgery, involving the patient's surrogate may be an alternative.

As in any case, the particular facts relevant to the situation will determine whether the removal of the tumor is covered under the scope of the informed consent. The facts will help determine whether the removal of the tumor is necessary to complete the surgery and what the risks and benefits associated with the removal of the tumor will be. After the surgery, the surgeons should inform the patient about the tumor and their reasons for either removing the tumor or waiting for the patient to make a decision.

Disclosure of Medical Errors

Most cases involving medical errors presented by the medical students during their surgery rotation, like the following one, focus on the question of disclosing the error to the patient. Medical errors encompass iatrogenic injury and errors in diagnosis, erroneous administration of drugs and other medications, technical errors during surgical procedures, and errors in interpretation of laboratory findings. Often doctors tie the need for disclosure of medical errors to the conspicuousness of the medical error. In cases in which the harm to the patient is likely to produce long-term or easily discoverable harm, doctors are compelled to disclose. In cases in which the error is not likely to significantly alter the patient's course, doctors justify nondisclosure by citing lack of harm. Although the decision to disclose errors must be made in view of the particularities of each case, in general, disclosure of medical errors is ethically required and is also good hospital policy.^{3,4}

Case

A 35-year-old female with biliary colic was scheduled for a laparoscopic cholecystectomy. Before the procedure, the patient signed a consent form on which were listed the possible complications of

bleeding, infection, and injury to the gallbladder and other internal organs and the possibility of conversion to an open procedure. During the case, the gallbladder was accidentally filleted open. Bile, more than 10 large gallstones, and countless small stones spilled into the peritoneal cavity. After removal of the gallbladder, the peritoneal cavity was repeatedly irrigated, and all visible stones were removed through port sites, although fragments were known to remain. The case was not converted to an open procedure. The patient received the appropriate antibiotic intraoperatively after the bile spillage. Postoperatively, the patient was told that everything went fine and she could be discharged that same day. There was no mention of the bile and gallstone spillage to the patient. Although most prospective studies in the surgical literature conclude that perforation of the gallbladder intraoperatively rarely causes complications, there does seem to be a very small increase in risk of abscesses and wound infections resulting from unretrieved gallstones.

Discussion

The question in this case is whether the surgeon has an ethical obligation to inform the patient that an error occurred during her surgery. Several concepts of medical ethics are involved: minimization of harm, respect for autonomy, and fiduciary responsibility.

The duty to minimize harm is sometimes offered as a justification for the decision not to inform a patient of a medical error such as the one that occurred in this case. Some doctors maintain that explaining the error will cause the patient unnecessary anxiety and fear that could, in turn, lead to detrimental health outcomes. This line of reasoning can justify withholding information from the patient in rare circumstances where the harm of disclosing the error will significantly outweigh the benefit for the patient.

Explaining that gallbladder contents were spilled into her abdominal cavity might cause the patient to worry that this error will negatively affect her recovery or subsequent health. Thus, the question becomes whether refraining from disclosing the error to the patient will prevent some clear harm without any important benefits being sacrificed. There is no reason to think that the harm being prevented is anything but a short-lived, psychologically normal anxiety response. In addition, there is reason to believe that the spill could negatively affect the patient's recovery and health, as indicated by the surgeon's decision to provide an antibiotic infusion intraoperatively.

Respect for autonomy mandates that physicians do more than simply honor their patients' decisions; it also obligates physicians to promote autonomy by informing their patients about factors relevant to their decisions. The occurrence of a medical error may well factor into a patient's subsequent decisions about what, if anything, to do about the medical error, how the results of the error should be monitored or treated, and even where to receive healthcare in the future.

Physicians are given a great deal of trust and control by their patients. This trust and control allow physicians to provide the best care possible for their patients. Yet, the trust and control that patients grant to their physicians comes with added obligations to honor that trust, which are often called the fiduciary responsibilities of medical professionals. One way in which a physician's fiduciary responsibility is relevant to the question of informing a patient about a medical error is by obligating a physician to put the patient's interest before his or her own. To use the trust and control ceded by a patient in order to further only the physician's interests would be a violation of the physician's fiduciary responsibilities. Thus, a physician may not conceal or refrain from disclosing a medical error in hopes of avoiding a lawsuit or unpleasant emotions and embarrassment. Therefore, during the postoperative meeting with the patient, the surgeon has an obligation to tell the patient the truth about the medical error and its possible consequences.

It is also of no small significance that this case was presented by a medical student who was herself involved in a complex triadic relationship with the patient and her supervising surgery team. This factor provides an additional reason to disclose the medical error to the patient, namely, to provide a learning experience for the student. Such a teaching opportunity is exceedingly valuable because it allows for the opportunity to model professional behavior.

Learning on Patients

Case

A 19-year-old woman arrives to the emergency room with a 3-cm laceration of her external ear. It is determined that the laceration is sufficiently severe to require surgery. During the informed consent discussion with the attending surgeon, the young woman makes it clear that she will consent to the surgery only if no residents or medical students participate in the procedure. She declares that she wants only "somebody experienced" to operate on her and "no students." The attending surgeon agrees and asks the residents and student to leave the examining room. During the surgery, however,

she goes about business as usual, with residents performing many parts of the surgery.

Discussion

Obviously, this case involves the physician's duty to tell the truth, particularly with respect to the role of learners who are involved in the patient's care. The related issue of whether patients may be used in medical education is also critical to this discussion. We shall begin with that issue first.

There is no evidence that the participation of medical students and residents in the healthcare team causes any harm to the patient. In fact, numerous studies have demonstrated that patients actually do better in teaching hospitals. For example, a recent comprehensive study comparing 132 academic and nonacademic healthcare structures found no significant difference in care between the 2 kinds of institutions.⁵

Nevertheless, we can formulate the ethical conflict in this way: it is in the patient's best interest to receive care only from the most highly skilled, well-trained, and experienced person available, who in this case is the attending surgeon. To allow anyone with less skill and less training to perform any piece of the surgery is not in the patient's best interest. At the same time, medical students and residents have a professional duty to develop their knowledge and skills, and hands-on training is necessary for trainees to attain the requisite level of expertise.

Anti-Free-Rider Argument No senior physicians working in academic teaching hospitals or in community hospitals attained their level of expertise solely through observing other physicians and reading books. Rather, they learned by doing. The attending in this (and every) case was once a resident and before that a medical student. She attained her level of expertise by studying under some senior physician before her and is now passing on that knowledge to the current resident. The patients before this one, who consented to being operated upon by the now attending when she was a still a resident, are in part responsible for the level of expertise that the attending has attained. It is, therefore, not fair for this patient to reap the benefits of the previous patients' consent to being operated on by residents without doing her fair share and helping to train the experts of tomorrow. Reaping the benefits of the contributions of others without contributing in kind is being a free rider on the system.

Golden Rule Argument Patients who demand to be treated only by the most senior physician

available are not morally situated differently than anyone else. In general, it is hard to imagine that anything about a particular patient would justify him being an exception to the rule of how others should be treated. If we accept the idea that others should participate in the education of tomorrow's doctors, we should participate ourselves.

In effect, the patient in our case is saying "I will be operated on only by someone who is already an expert at what he does." If everyone acted on that principle, then there would be no expert surgeons. No one would consent to surgery by residents, and then no surgeon would gain the requisite experience in order to become an expert. No one could consistently choose that state of affairs, and our patient could not consistently choose that everyone would refuse to participate in medical training as she would like to do. Taken together, these arguments imply that this patient should not demand treatment by only the most senior physician to the detriment of the professional development of future physicians. Nevertheless, she does, and the Patients' Bill of Rights allows the patient to choose her doctors. At the same time, the attending surgeon's duty to medical trainees is to aid in their professional growth and development of competence. The duty of residents and students is to become competent physicians. Clearly, the patient's right clashes with the duties of the attending surgeon and trainees.

Often, patients are not aware that by seeking treatment in teaching hospital, they are likely to be cared for by physicians in training. It is imperative that senior physicians inform their patients about the involvement of learners on the surgical team. They need to explain who the various learners are and what their roles are on the team. It is also important for residents and students to clearly identify themselves and their role and position on the medical team to patients.

ETHICS PROJECT IN THE SURGERY CLERKSHIP: SURVEY RESULTS

Student feedback in an end-of-clerkship evaluation suggests that the ethics project serves an important function in the third-year surgery clerkship and contributes to medical student ethical development. The ethics project has enjoyed strong support from the student body over the past 2 academic years, with 85% of students answering yes when asked if the ethics project was useful. The survey of student reactions to the project provides insights into the impact of clinically based medical ethics education

on the experience of students first engaging with the surgical field.

When asked to comment on the utility of the ethics project, students frequently remarked that the exercise raised their awareness of ethical dilemmas within the surgery clerkship. One student noted that "it made me keenly aware of the issues that I was constantly surrounded by on the wards." Another commented that "in ASM 1 and 2, the cases presented were straightforward. Out in the hospital, the cases were somewhat more subtle. Since they were pulled from real-life situations that may occur to me in the future, the discussions really challenged me to think about how to act with ethical integrity." Several students commented that they had never before considered certain issues such as health insurance coverage to be ethics topics until presented by their peers.

The ethics project also serves as a source of psychosocial support for the class of surgical clerks. By coming together to discuss core issues of medical ethics, the group process solidifies the collective consensus around ethical issues that were introduced during the first and second years. One student remarked that "it was great to sit with a classmate, discuss these issues, and learn from them." Another student wrote that the act of documenting a troubling experience through the narrative of the ethics project was valuable. Another student put the session in context, commenting that the project encouraged "reflection upon experiences that were disturbing while giving us a venue to productively discuss these experiences." By providing a model for clinical reflection among peers, the ethics project suggests a remedy for feelings of isolation and even moral distress experienced by many young clinicians. The project contributes to the creation of a community of mature ethical thinkers who understand the importance of drawing support from one another.

Student criticism of the ethics project was overwhelmingly related to logistical issues. A few students expressed cynicism in their evaluation of the project, wondering if discussions ever changed standards of practice and widespread views on ethical behavior. Generally, students were in agreement that the interdisciplinary approach was helpful and contributed to the discussion.

Although the objective of the ethics component of the clerkship is primarily medical student education, its design also creates opportunity for reflection on the part of the surgeon preceptors who participate in the sessions. These preceptors, although untrained in the language of ethics, are the sole representatives of the culture of surgery.

As such, they describe feelings of humility and embarrassment when listening to the anecdotes presented by students, which often detail unethical behavior on the part of their colleagues. One resident wrote that her participation in the sessions “reminded her why we all become doctors . . . we tend to forget and get distracted by so many other things during residency.” That doctors normalize their daily routine in order to maintain sanity in the face of the weighty responsibility of human lives is not surprising; however, students often perceive this normalization as bordering on unethical.

CONCLUSION

The tabulation of students’ ethics project topics reveals that particular ethical dilemmas tend to recur in the practice of surgery. It reveals that the students were competent and comfortable discussing ethical issues and also suggests that the conceptual background in basic principles of medical ethics introduced during the Arts and Science in Medicine course was successful in helping them learn to identify ethical issues and to analyze them in terms of moral conflicts.

The clerkship survey results demonstrate the value of conducting ethics sessions in the third- and fourth-year clerkships focusing on those dilemmas that arise most vividly in each specialty. The feedback on the sessions was overwhelmingly positive. The sessions provided an opportunity for students to express their views on the ethical issues involved in the cases presented for discussion. By discussing the ethical dilemmas with fellow students, surgeons, and members of the ethics faculty, presenters felt that their ethical concerns were legitimized. Furthermore, the ethics sessions provided an opportunity for students to observe the surgeons who were teaching them in a way that did not just focus on the acquisition of medical skill. Students were able to focus on their

own future identity as physicians by assessing the ethics of their teacher’s behavior, and this enabled them to decide what type of physician they aspired to become.

The study of this surgery clerkship activity demonstrates that ethics education integrated into a 4-year program of medical education can develop critical sensitivity to ethical issues and prepare trainees to address the ethical issues that they will encounter in clinical practice. It also provides a model for how to do the job well. Furthermore, the themes identified as ethical issues in surgery could be used to develop educational activities for surgical house staff and even surgical faculty in order to allow these common issues to be preemptively addressed.

ACKNOWLEDGMENT

The authors thank Shariff Ellozy for his involvement in this project.

DISCLOSURES

Potential conflict of interest: No conflicts of interest. Nothing to report.

REFERENCES

1. Rhodes R, Alfandre D. A systematic approach to clinical moral reasoning. *Clin Ethics* 2007; 2: 66–70.
2. Dworkin G. Paternalism. *Monist* 1972; 56: 64–84.
3. Brennan TA. Ethics of disclosure following a medical injury: time for reform. In: Rhodes R, Francis LP, Silvers A, eds. *The Blackwell Guide to Medical Ethics*. Oxford, England: Blackwell; 2006; 393–406.
4. Sack K. Doctors say ‘I’m sorry’ before ‘See you in court.’ *New York Times* May 18, 2008.
5. Papanikolaou PN, Christidi GD, Ioannidis JPA. Patient outcomes with teaching versus nonteaching healthcare: a systematic review. *PLoS Med* 2006; 3: e341.

SPECIAL FEATURE

Testing for Thrombophilia: Pitfalls, Limitations, and Marginal Impact on Treatment Duration Recommendations

Richard H. White, MD,¹ and Robert C. Gosselin, CLS²

¹Department of Internal Medicine, School of Medicine, University of California Davis, Sacramento, CA

²Department of Pathology and Laboratory Medicine, School of Medicine, University of California Davis, Sacramento, CA

A great deal of information has been published in the past 10 years about both inherited and acquired humoral or cellular risk factors that are associated with the development of venous thromboembolism (VTE). These disorders are associated with hypercoagulability and are collectively called thrombophilic disorders.^{1,2} This research has been very exciting, and many clinicians can now order a whole panel of these tests at the time they diagnose a patient as having either deep vein thrombosis or pulmonary embolism.³ Although there may be a certain sense of satisfaction associated with identifying and then labeling a patient as having one of these specific thrombophilic conditions, it is not clear that the patient benefits in any way or that the management of the patient changes. Although there may be some benefit to testing for thrombophilia, there are many pitfalls and perils associated with testing that may be detrimental to either the patient or his or her family. The most common thrombophilic conditions are reviewed here (see Table 1), with an emphasis on the complexity of performing and interpreting the laboratory tests for these disorders and how testing affects patient management. The issue of thrombophilia testing in women who have a history of complications of pregnancy, such as pre-eclampsia, is not discussed.

MAJOR THROMBOPHILIC DISORDERS

The thrombophilic condition most commonly recognized by primary care physicians is the inherited trait called factor V Leiden (FVL). However, it is instructive to begin a discussion of these disorders by starting with the first conditions that were fully characterized and described. The first hypercoagulable disorders that were fully characterized were anti-thrombin (AT) deficiency⁴ and partial deficiency of protein C (PC) and protein S (PS).⁵ These disorders were the first to be elucidated because each of these proteins was known to be important in the regulation of blood coagulation and each could be readily measured in large numbers of patients with VTE and in their family members. A great deal of work by many investigators showed that the number of genetic variants that lead to abnormal function of AT, PC, and PS is both large and heterogeneous,^{6–8} with mutations that can result in quantitative dysfunction (decreased production; typically called type I deficiencies) or qualitative dysfunction (abnormal function; typically called type II deficiencies).^{1,9,10} Given this vast heterogeneity, it is likely that patients with different genetic disorders have significantly different levels of risk for thrombosis;¹¹ nevertheless, patients are currently labeled and grouped into single disorders, such as PS deficiency.

Laboratory measurements of coagulation proteins can assay either the functional component of the protein or the amount of protein present (antigenic measurements). As a general rule, the functional measurements of coagulation proteins use either a clotting or chromogenic endpoint. Clotting endpoint assays can be problematic because

Address Correspondence to:

Richard H. White

Division of General Medicine
University of California Davis
Sacramento, CA
Email: rhwhite@ucdavis.edu

Table 1. *Most Common Thrombophilic Disorders.*

Thrombophilic Condition	Prevalence (Population)	Inherited or Acquired
Factor V Leiden	3%–7%	Inherited
Prothrombin G20210A	1%–2%	Inherited
Protein C deficiency (PC)	0.3%	Inherited
Protein S deficiency (PS)	0.1%	Inherited
Anti-thrombin deficiency	0.1%	Inherited and acquired
Elevated factor VIII	10%	Acquired
Anti-phospholipid antibody syndrome	1%	Acquired
Hyperhomocysteinemia	5%	Acquired
Activated protein C (aPC) resistance without factor V Leiden	0.5%	Inherited and acquired

many variables can influence in vitro clot formation (eg, factor deficiencies, inhibitors, and drugs), and this may result in false reports of elevated or depressed protein function. Chromogenic assays tend to have less interference but are more costly than clot-based methods. Antigenic determination of the levels of proteins is commonly performed with enzyme-linked immunosorbent assay–based testing, but other immunoassays (eg, electrophoresis and latex immunoassays) can also be used. Most antigenic assays do not have the sources of interference associated with functional assays, but performing only antigenic assays may result in misdiagnosing patients who have a type I (functional) protein deficiency.

Anti-Thrombin Deficiency

In adults, AT is the principal molecule that inhibits thrombin and other serine proteases, such as activated factor X. AT is the molecule to which heparin binds, and this in turn leads to irreversible inhibition of thrombin by the binding of AT to the active site of thrombin. AT deficiency is an autosomal dominant disease, with homozygous type I AT deficiency not compatible with fetal survival. The exact prevalence of heterozygous AT deficiency is not known, but it is estimated to be present in perhaps 0.02% of the population.¹² Importantly, therapeutic levels of heparin lower the level of AT up to 20% because of direct binding to AT.¹³

Low levels of AT may also reflect an acquired condition, which is frequently seen in patients who have serious medical conditions such as liver disease, sepsis, polytrauma, or consumptive coagulopathies.¹⁴ In such patients with acquired AT deficiency, high levels of heparin may be required to prolong the activated partial thromboplastin (aPTT), leading to so-called heparin resistance. Rare

qualitative defects of the AT-heparin binding have also been described. In the laboratory, AT function is most commonly measured with a chromogenic-based assay. With this method, the AT present in a patient's plasma neutralizes exogenously added thrombin or activated factor Xa. In addition to the aforementioned illnesses and heparin therapy, falsely depressed levels of AT may occur in patients receiving direct thrombin inhibitors (when thrombin-based AT measurements are used) and in patients with marked bilirubinemia or lipemia (interference with color determination). Antigenic determination of AT is rarely performed but is available through specialized reference laboratories.

Protein C and Protein S Deficiency

Both PC and PS are important in the inactivation of activated coagulation factors Va and VIIIa. Thus, in addition to AT, PS and PC are the key proteins in the body's natural anticoagulant system. A reduction in the level or function of these proteins increases the likelihood of thrombosis. Like patients with AT deficiency, patients may have inherited deficiency in the level of functionally normal PC or PS, or they may have normal levels of a dysfunctional protein.⁸ Multiple mutations (>150) have been described for each disorder.¹⁵ Homozygous PC or PS deficiency presents with neonatal purpura fulminans or cerebral vein thrombosis.

Initial reports suggested that PC deficiency leads to a high risk of VTE. However, a study reported by Miletich *et al.*¹⁶ caused great confusion because a large number of blood donors and family members who had PC deficiency had no reported increase in the incidence of venous thrombosis. This finding suggested either incomplete genetic penetrance, with only some mutations leading to thrombotic risk, or that a secondary predisposition was needed to result in thrombosis. The explanation of this conundrum was elucidated some years later when it was discovered that the patients and families who manifested thrombosis had PC deficiency together with a second condition called resistance to activated protein C (aPC), which was later identified as being caused by a mutation on factor V (FVL).¹⁷ This finding led to the currently popular view that most patients who develop thromboembolism likely have a) 2 underlying prothrombotic conditions (such as PC plus FVL), b) 1 transient provoking risk factor combined with 1 thrombophilic disorder (such as a leg fracture plus a thrombophilic disorder), or c) a single very strong provoking risk factor (such as cancer).

Interestingly, other studies have also reported that individual carriers of PS deficiency do not have a

significantly increased risk of developing VTE.¹⁸ This may be in part due to the way in which PS travels in circulation or is measured in the laboratory. In vivo, PS is either bound or free. Bound PS is the amount of protein that is bound to the negative acute-phase reactant, C4b, and is not functional.¹⁵ The free portion of PS is unbound to C4b and constitutes the functional portion of PS. Therefore, conditions that result in elevated C4b (eg, infection and pregnancy) can lead to decreased PS activity. Laboratory methods used to measure PS include measures of the total antigen (bound plus free), and free PS and a functional PS method that depends on the free portion of PS. PS function is always measured with clot-based testing, as there are no chromogenic methods are currently available. Functional PS is measured with a modified aPTT, with PS prolonging the time needed to clot.¹⁹ PS activity may be influenced by the presence of FVL (with some methods)¹⁹ and by pregnancy and estrogens.

In most laboratories, the functional level of PC is measured with either a chromogenic assay or a clot-based assay.¹⁵ The desired method is the chromogenic assay because less interference is noted with this methodology. Clot-based PC determinations typically use a modified aPTT, and high levels of factor VIII or fibrinogen can result in falsely reduced PC levels, whereas heparin, direct thrombin inhibitors, and lupus anticoagulants (LAs) can falsely increase the reported PC activity.

It is unusual for a patient to have both PC and PS deficiencies. These 2 conditions are so rare that finding a patient with both is prohibitively unlikely. However, because the activity of both proteins requires vitamin K–dependent carboxylation, when the levels of both proteins are low, it is overwhelmingly likely that this reflects the fact that the patient was taking warfarin at the time of testing. In addition to the causes previously described, other causes of falsely depressed levels of PC and/or PS include the acute phase of venous thrombosis, consumptive coagulopathies, and liver disease. Because there are numerous problems associated with PC and PS testing in the acute setting of thrombosis, it is prudent to confirm deficiencies of these proteins 2 to 3 weeks after completion of anticoagulant therapy.

Factor V Leiden

First described as the principal cause of aPC resistance, FVL is a point mutation in the factor V molecule that leads to a reduction in the efficiency of the anticoagulant properties of aPC, which can also be caused by other genetic defects or acquired

conditions.¹² Although the majority of aPC-resistance patients have the V Leiden mutation, other point mutations on factor V that cause aPC resistance have also been described. Patients with FVL have a slightly greater predilection to form or propagate thrombi. It is the most common genetic disorder known, as approximately 5% to 7% of Caucasians from Northern Europe are affected. Identification of FVL can be performed at any time with a polymerase chain reaction test. It must be stressed that the presence of FVL does not constitute a disease per se; rather, it is a genetic variant that is associated with a modest 3- to 4-fold increase in the lifetime risk of developing VTE in comparison with individuals without this condition.²⁰ However, the absolute risk of developing VTE is rather low, with 80% of individuals never developing VTE by age 90.²¹ A large percentage of women who develop VTE in the first year or two after starting oral contraceptives have been shown to have FVL (high relative risk). However, only a very small percentage of women with FVL who take estrogen-containing oral contraceptive pills ever develop VTE (low absolute risk).²²

Factor II Mutation: The G20210A Prothrombin Gene Mutation

The second most common genetic disorder is the prothrombin gene mutation (G20210A), a point mutation that is associated with increased levels of factor II, or prothrombin. Exactly why this mutation leads to higher levels of prothrombin and a higher risk of thrombosis is not known. Identification of the G20210A mutation can be performed at any time with a polymerase chain reaction test. Factor II (prothrombin) levels are not an appropriate test to order because there is an overlap between normal reference values and the prothrombin activity levels found in patients with the G20210A mutation.²³ The risk of thrombosis appears to be modestly increased, being approximately 2- to 3-fold higher than the risk for unaffected individuals, which is slightly lower than the risk associated with FVL. The prevalence of this trait is increased 10-fold in individuals with cerebral venous thrombosis, and use of oral contraceptives increases the risk of developing VTE. The risk of VTE increases in a multiplicative fashion when both G20210A and FVL are present.

Anti-Phospholipid Antibody Syndrome

The criteria for classifying a patient as having anti-phospholipid antibody syndrome (APS) are shown in Table 2.²⁴ This is an acquired autoimmune condition with many features and laboratory findings that overlap with systemic lupus erythematosus. Diagnostic

Table 2. *Criteria for Classification of Patients with Anti-Phospholipid Antibody Syndrome.*

Clinical Criteria (At Least 1)	Laboratory Criteria (At Least 1)	Level*
1) Venous or arterial thrombosis	1) Lupus anticoagulant	
2) Pregnancy morbidity	2) Anti-cardiolipin antibody	>40 U or 99th percentile
a) Unexplained fetal loss at >10 weeks OR	3) Anti-beta-2-glycoprotein I	>99th percentile
b) Premature delivery with eclampsia, pre-eclampsia, or placental insufficiency OR		
c) Three or more unexplained consecutive spontaneous abortions at >10 weeks		

NOTE: Anti-phospholipid antibody syndrome is subclassified as follows: I, more than 1 laboratory criterion present (any combination); IIa, only lupus anticoagulant present; IIb, only anti-cardiolipin antibody present; and IIc, only anti-beta-2-glycoprotein I antibody present.

*Repeatedly positive at least 12 weeks apart.

criteria are stringent and not widely appreciated,²⁵ and this leads to frequent overdiagnosis. The long-term risk of recurrent venous or arterial thrombosis in patients with APS is not well defined. Nevertheless, patients with an acute thrombosis who meet laboratory criteria for APS generally are treated indefinitely with warfarin, with a target international normalized ratio of 2.0 to 3.0.^{26,27} The 3 laboratory tests that may be used to fulfill the laboratory criteria for APS include anti-cardiolipin (aCL) antibody, anti-beta-2-glycoprotein I (anti-beta-2GPI) antibody, and the lupus anticoagulant (LA). To meet the criteria, elevated levels of immunoglobulin G or immunoglobulin M aCL antibodies (>40 GPL or MPL units, respectively), elevated levels of immunoglobulin G or immunoglobulin M beta-2GPI antibodies (>20 GPL units or >20 MPL units, respectively), and/or an LA ratio of >1.2 must be found, and they must remain elevated when the test is repeated at least 12 weeks later. Initial testing for APS is most appropriate shortly before initial anticoagulation therapy is planned to be discontinued, generally 12 or 24 weeks after the index event. During oral anticoagulation treatment, aCL and anti-beta-2GPI should be measured. If either of these tests meets the criteria for APS, anticoagulation should be continued until repeat testing is performed 12 weeks later.

Although there are fairly strict guidelines for identifying patients with APS, the testing used for identifying antibodies to phospholipids is less stringent. The majority of laboratories use either commercially prepared or home-brew enzyme-linked immunosorbent assay methods. In a recent US survey of laboratories performing aCL testing, when 1 specific sample was measured in all the participating laboratories, the mean aCL level ranged from 17 to 121 units (mean of methods: 24–79 MPL units).²⁸ There are no accepted standardized tests or accepted standards for measuring aCL or beta-2GPI, and this leads to a great deal of variation in reported results.

Testing for LA is even less standardized. The principle of LA testing is to determine if there is an antibody present in the plasma that interferes with the thromboplastin reagent. With the exception of the dilute Russell viper venom time (dRVVT), most LA testing is based on a modified aPTT. Changing the thromboplastin used in the aPTT or the use of either platelets or platelet components may improve the sensitivity and specificity of the LA test. One aPTT reagent, Actin FS (Dade Behring, Newark, DE), does not detect LA at all. Ideally, in a patient who has the LA, an elevation of the aPTT is observed that is due to the LA (antibody) binding to the phospholipids contained in the reagent, thereby prolonging the time required for clot formation. Once a screening aPTT is noted to be prolonged, a secondary aPTT containing an increased (excess) amount of target phospholipid is then used. A correction of the clotting time toward normal is seen after the addition of the extra phospholipid. This second step should document that the initially prolonged aPTT was due to a phospholipid-specific antibody rather than a different cause (eg, factor deficiency, drug effect, or inhibitor of factor VIII).

The dRVVT is another means of measuring an LA, and it is most reliable when performed in patients who are not receiving any anticoagulants. Although the dRVVT is not the same as an aPTT (it directly activates factor X), the principle of the test is the same as that for other LA tests: an initial screening test is used, that contains low levels of phospholipid, and if this clotting time is prolonged, a second dRVVT test is performed with increased amounts of phospholipid. If the addition of phospholipid corrects the clotting time, this indicates that there is a phospholipid-dependent antibody present. All the different tests that are available differ in their sensitivity and specificity for detecting LA, especially low levels of LA. In a recent survey in the United States and Europe, there was a wide discrepancy in detecting

a weak LA with a 1 test sample, with approximately 70% of the laboratories that used aPTT-based assays reporting negative results, whereas only 45% of the laboratories that used the dRVVT tests reported the sample as LA-negative.²⁹ Although there is no gold-standard test for detecting LA, dRVVT testing has been suggested as the best standard test for LA.

The use of laboratory kits that measure antibodies to other phospholipid antibodies has not yet gained acceptance. These include the generic anti-phospholipid antibody test (anti-phospholipid cocktail target) as well as antibodies directed to phosphatidyl serine and anti-prothrombin antibodies. Of note, there have been reports of LA patients with high levels of anti-prothrombin antibodies with rare bleeding diathesis.²⁹ Additionally, there is mounting evidence that aCL testing may not be sufficiently sensitive and specific, so aCL testing may be deleted in future APS classification guidelines.

OTHER THROMBOPHILIC DISORDERS

Other conditions classified as being thrombophilic conditions include elevated factor VIII (>150 IU/mL, 150%), very elevated levels of homocysteine during fasting (top 10%), dysfibrinogenemia, and other uncommon conditions.^{1,12,20,30,31} Because factor VIII is an acute-phase reactant, the identification of patients who have persistent elevations of factor VIII requires repeat measurements after the patients recover to their baseline state of health.³² Patients with homocysteine levels above the 90th percentile have been shown to have a 2- to 4-fold increased risk of developing VTE.^{30,31} A recent controlled trial that evaluated the effect of treatment using folic acid and B vitamins dampened the interest in identifying patients with elevated homocysteine. This supplementation led to a significant fall in the homocysteine level, but there was no change in the incidence of VTE.³³ Heterozygous carriers of the thermolabile variant of methylenetetrahydrofolate reductase do not appear to be at increased risk for VTE, and this gene polymorphism is not strongly associated with homocysteine levels, which correlate more closely with folic acid intake.³⁴

DOUBLE OR COMBINED THROMBOPHILIC DEFECTS

There is a large body of evidence suggesting that patients with 2 thrombophilic defects have a significantly higher relative risk of developing

incident or recurrent VTE in comparison with patients with simply 1 defect.^{35,36} This is illustrated by the discovery that kindreds with PC deficiency which manifested symptomatic VTE also had FVL.¹⁷ It is also known that the risk of venous thrombosis is much higher in individual patients with homozygous FVL or homozygous G20210A in comparison with heterozygotes and that patients with FVL plus the G20210A mutation have a higher incidence of VTE in comparison with patients with either defect alone.³⁷⁻³⁹ The fact that patients with more than 1 thrombophilic defect appear to be at a higher risk for incident and recurrent VTE is an argument in favor of thoroughly testing all patients with VTE for all of the known thrombophilic conditions.⁴⁰

In summary, some of the pitfalls in testing for thrombophilia are as follows: acute illness or initial anticoagulation can affect the measurement of AT levels (by heparin), proteins C and S (illness and inflammation), factor VIII (illness and inflammation), aCL antibodies (high levels associated with some infections such as syphilis), and the dRVVT (higher on heparin or warfarin). FVL may depress the measured level of PS, and anti-phospholipid antibodies can affect clot-based assays for PC and PS or depress PS *in vivo*.⁴¹

RISK FACTORS FOR VENOUS THROMBOEMBOLISM AND RECURRENT VENOUS THROMBOEMBOLISM

Incident Venous Thromboembolism

Table 3 shows the most common demographic and clinical features associated with an increase in the risk of first-time VTE or recurrent VTE. Disorders such as recently diagnosed metastatic cancer, trauma, and major surgery are potent provoking risk factors associated with a very high incidence of both asymptomatic VTE and symptomatic VTE. Some studies show that incident VTE is more common in patients with a provoking risk factor who also have a thrombophilic disorder, but the added risk is too small to warrant screening for the presence of these disorders.^{22,41} Other factors such as obesity, presence of chronic venous stasis, older age, immobility, use of oral contraceptives, and the presence of an acute medical illness such as sepsis or active lupus erythematosus are clearly risk factors, but the relative strength of each of these risk factors has not been well defined.⁴² Overall, 40% to 70% of patients who are diagnosed with VTE have 1 of the common thrombophilic disorders.⁴³⁻⁴⁵ A small

Table 3. Risk Factors for Incident and Recurrent Venous Thromboembolism.

Incident Venous Thromboembolism	Recurrent Venous Thromboembolism
Advancing age	Unprovoked incident venous thromboembolism
Exposure to a provoking risk factor, especially cancer, trauma, major surgery, and immobilization	Male sex > Female sex
Obesity	Older age
Race/ethnicity	Race/ethnicity
Chronic comorbidity (eg, lupus or Crohn's disease)	Presence of postphlebotic signs or symptoms
Acute illness (eg, sepsis)	Obesity
Inherited or acquired thrombophilia	Immobility
Specific drugs chemotherapy	Persistently elevated d-dimer
Oral contraceptive or hormone replacement therapy	Adherence to initial treatment
	Failure of resolution of thrombosis with ultrasound
	Placement of an inferior vena cava filter

fraction of patients have 2 thrombophilic disorders. There may be synergism between thrombophilic disorders and environmental risk factors, and this has been documented most convincingly in women with FVL who are using oral contraceptives.

Recurrent Venous Thromboembolism

It is widely accepted that patients who have a VTE event after exposure to a significant transient risk factor, often termed a provoking or secondary risk factor, have a lower incidence of recurrent VTE in comparison with patients with an unprovoked or idiopathic VTE event.⁴⁰ It is also widely acknowledged that women with unprovoked VTE have a 2- to 4-fold lower incidence of recurrent VTE in comparison with men,⁴⁷ even after the specific exclusion of women taking oral contraceptives or hormone replacement therapy.⁴⁸ Importantly, a number of studies have shown that the presence of a single thrombophilic disorder is not associated with a significant difference in the risk of a recurrent VTE event in patients with unprovoked VTE.⁴⁹ However, it should be acknowledged that these studies excluded patients with known AT deficiency, APS, or known PS or PC deficiency.

The most likely reason that patients with a thrombophilic disorder do not have a higher incidence of recurrent VTE in comparison with individuals without a thrombophilic disorder is that the patients labeled as having no thrombophilia do in fact have an underlying procoagulant condition; it simply has not yet been characterized. Imagine 2 identical (weight, race, and sex) and otherwise healthy patients who, for example, are 30 years old and present with an acute unprovoked proximal deep vein thrombosis, with no other underlying illnesses. After testing, one is found to have only FVL, and the other has no thrombophilic condition. Clearly, the patient without FVL has something that predisposed him or her to thrombosis at such a young age; we just do not know what it is. This is why screening for thrombophilia usually does not affect patient management: absence of thrombophilia does not predict a better long-term outcome.

Another critically important point that needs to be stressed is that most studies that have analyzed the association between specific thrombophilic defects and recurrent VTE have not taken into consideration the myriad of other patient-specific risk factors that may be much more important in predicting recurrent VTE.^{37,38,45,49–57} These studies have, in general, analyzed only the effect of the presence of thrombophilia. Other important risk factors that likely play a critical role include age, immobility/frailty, severe obesity, the presence of chronic venous insufficiency, and the presence of comorbidity together with adherence to the initial course of oral anticoagulation therapy. Other factors that may also play a pivotal role in leading to recurrent VTE include the degree of resolution of the thrombus and evidence of ongoing activation of the coagulation system from d-dimer levels either during warfarin therapy or after the conclusion of warfarin therapy.^{53,54} The Recurrent Venous Thromboembolism Risk Stratification Evaluation Study (REVERSE) analyzed many predictors of recurrent venous thrombosis and found that thrombophilia was not a significant predictor of recurrent VTE.⁵⁸

TESTING FOR THROMBOPHILIA

When a patient is admitted with a documented pulmonary embolism or acute deep vein thrombosis, the following question is often raised: should the patient be worked up for a hypercoagulable state? This is a highly contentious issue,⁵⁹ with many experts divided, some favoring testing of essentially all patients with an idiopathic venous

Table 4. *Arguments For and Against Thrombophilia Testing.*

Arguments in Favor of Testing	Arguments Against Testing
Testing provides the opportunity to define the cause of thrombosis.	Half of all venous thromboembolism patients with venous thromboembolism will not have a known thrombophilic disorder. These patients without a known thrombophilic disorder probably do have a thrombophilic disorder that has simply not been discovered yet. Finding thrombophilia does not affect any treatment decision. Costs of testing are significant.
Results influence treatment decisions; there is a greater risk of recurrent thrombosis.	There is no evidence that the presence or absence of a single thrombophilic defect affects the risk of recurrent thrombosis. Different doctors may interpret results differently, creating confusion and anxiety. Studies that show modestly higher risk in patients with a thrombophilic disorder have not been risk-adjusted for other major risk factors.
Patients with venous thromboembolism who have thrombophilia should receive more aggressive prophylaxis in the future.	Results should not affect future prophylaxis or treatment. All patients with previous venous thromboembolism are at increased risk and need more aggressive prophylaxis.
Testing is useful for screening family members of an affected proband with thrombophilia; documenting the presence of thrombophilia may affect future treatments.	There is no evidence that screening family members affects the subsequent risk of thrombosis in only the family members who test positive. One should treat family members as potentially having thrombophilia, regardless of test results. Testing may create great anxiety in unaffected family members.
Testing may detect high-risk patients with 2 or more defects.	Clinical features combined with a family history are more useful for defining high-risk patients.
Testing is useful in managing patients with previous venous thromboembolism who are considering pregnancy.	Management should be the same, regardless of the thrombophilia workup.
Testing is useful in managing women who want to become pregnant and who have a family member with known thrombophilia.	There is no evidence to support this view. However, if an asymptomatic family member is found to be homozygous for factor V Leiden or G20210A, prophylaxis should be offered, but this situation will be rarely encountered.
Patients with anti-thrombin deficiency may be resistant to heparin and have a higher recurrence rate.	Low levels may be acquired; heparin lowers the level of anti-thrombin, and heparin resistance will be clinically evident. When it is inherited, a positive family history is usually present.
Long-term treatment is indicated in patients with anti-phospholipid antibody syndrome.	Strict diagnostic criteria must be met! Results are frequently misinterpreted, and the patient is anticoagulated for life.

thrombotic event,^{20,60} some favoring no testing whatsoever,^{10,12,41} and some advocating selected testing⁶¹ (Tables 4 and 5). At the base of this controversy is disagreement regarding whether identification of a thrombophilic condition or a pair of conditions affects subsequent treatment decisions.⁵⁹

In a patient with an acute VTE, there is no need to immediately test for a thrombophilic disorder because the results of this testing will not alter clinical management, which usually is parenteral anticoagulation using heparin or low-molecular-weight heparin followed by oral anticoagulation with warfarin. Nevertheless, in a patient who manifests heparin resistance (failure to increase the aPTT), it may be reasonable to measure the functional level of AT. The patients most likely to have low levels of AT are young individuals with unprovoked VTE

who have a family history of VTE and acutely ill polytrauma patients. The clinician must recognize, however, that because heparin binds to AT, heparin itself lowers the level of functional AT by up to 20%.

It is reasonable to test every patient with idiopathic VTE for APS because all current guidelines recommend that APS patients should be maintained on oral anticoagulant therapy indefinitely.⁶² The 2 tests for APS that should be ordered because they are not affected by heparin or warfarin are the tests for aCL and anti- β -2GPI. Testing for the LA using the dRVVT can be performed, but it is best measured after the patient is off all anticoagulants. Because criteria for APS require repeat testing over time,²⁴ initial testing provides 1 set of values that will need to be repeated 12 weeks later. An initially positive test for aCL or β -2GPI antibodies does not confirm

Table 5. *Pitfalls in Testing for Thrombophilia.*

Methods	Specific Tests	Testing Pitfalls
Genetic tests	Factor V Leiden mutation Prothrombin G20210A gene mutation MTHFR gene mutation	Polymerase chain reaction assay can be performed at any time. It is not advised as a screening test prior to the use of birth control or pregnancy. It will not detect rare aPC-resistance patients who do not have the Leiden mutation. The MTHFR mutation does not correlate well with thrombotic risk.
Clot-based assays	PC activity PS activity	Factors that may decrease aPTT may falsely reduce activity levels (eg, factor V Leiden, increased fibrinogen, and FVIII). Factors that may increase aPTT may falsely increase activity levels (eg, high-dose heparin and direct thrombin inhibitors). PS activity decreases with acute inflammation (due to C4b binding) and pregnancy, and both PC and PS may be reduced by warfarin therapy, consumptive coagulopathies, trauma, sepsis, liver disease, systemic lupus erythematosus, APS, nephrotic syndrome, and other causes.
	Lupus anticoagulant (dRVVT test)	This test may be affected by anticoagulant therapy, including warfarin, heparin, and direct thrombin inhibitors. Other inhibitors and factor deficiencies may also influence the test. There is no standardized testing or recommended test method at this time.
	FVIII activity	This test may be influenced by heparin and direct thrombin inhibitor therapy.
	aPC resistance	This test may be influenced by heparin and direct thrombin inhibitor therapy.
Chromogenic assays	AT activity PC activity FVIII activity	This test may be influenced by heparin and direct thrombin inhibitor therapy. There is possible interference with the factor V Leiden mutation. It is more costly than clot-based methods. Its effectiveness may be reduced by warfarin therapy (PC only), consumptive coagulopathies, trauma, sepsis, liver disease (except FVIII), nephrotic syndrome, and other causes.
Antigen assays: ELISA	Total and free PS levels PC antigen level AT antigen level	The test is time-consuming. It has unclear utility for patients on warfarin therapy. It may miss functional (type II) deficiencies if only the antigen assay is performed. Its effectiveness may be reduced by warfarin therapy, consumptive coagulopathies, trauma, sepsis, liver disease, systemic lupus erythematosus, APS, nephrotic syndrome, and other causes.
Antibody assays: ELISA	aCL antibody	The test is considered positive or elevated only if there are more than 40 GPL or MPL units and the level remains elevated when repeat testing is performed 12 weeks later. There is no standardized or recommended testing method. False positive results may be seen because of seropositive syphilis, increased rheumatoid factor, and nonspecific drug interference.
	aPL antibody	Results have not been validated. It is not currently recommended. There is no standardized or recommended testing method. False positive results may be seen because of seropositive syphilis, increased rheumatoid factor, and nonspecific drug interference.
	Anti-beta-2GPI	The test is elevated if there are >20 GPL or >20 MPL units; the test must be repeated and elevated at least 12 weeks later. There is no standardized or recommended testing method. False positive results may be seen because of seropositive syphilis, high rheumatoid factor levels, and nonspecific drug interference.
Other	Homocysteine	Fasting is required; the specimen must be handled correctly.

Abbreviations: aCL, anti-cardiolipin; anti-beta-2GPI, anti-beta-2-glycoprotein I; aPC, activated protein C; aPL, anti-phospholipid; APS, anti-phospholipid antibody syndrome; aPTT, activated partial thromboplastin; AT, anti-thrombin; dRVVT, dilute Russell viper venom time; ELISA, enzyme-linked immunosorbent assay; FVIII, factor VIII; MTHFR, methylenetetrahydrofolate reductase; PC, protein C; PS, protein S.

a diagnosis of APS; rather, the result simply indicates that the patient needs repeat testing.

Comprehensive testing for thrombophilia, when performed, should be delayed until the end of the initial treatment period, which according to American College of Chest Physicians guidelines is 3 months for a provoked VTE and at least 3 to 6 months for an unprovoked VTE.^{63,64} The optimal time to consider testing for a thrombophilic disorder is the end of the initial treatment period. The key question that must be asked at this time is whether the results of the testing will affect the decision to continue or stop long-term oral anticoagulation.

In patients being treated for an unprovoked VTE, the decision to continue or stop anticoagulation after 3 to 6 months is often based on issues that have little to do with the etiology of the VTE. Some patients are adamant that they want to stop warfarin, even after being educated about the risks of recurrent VTE and after acknowledging that they understand the risks. Other patients are anxious about developing a recurrent clot and want to stay on warfarin indefinitely. As mentioned previously, testing for APS is appropriate because the findings do affect the duration of treatment, but the presence or absence of any other thrombophilic disorder does not affect the duration of treatment.^{45,58,64,65}

Routine testing for thrombophilia may have a number of unexpected consequences. For example, once a test result is in the chart, other physicians may incorrectly interpret the result. Although one physician may dismiss the importance of the finding of FVL, another may conclude that the patient has to stay on warfarin for the rest of his or her life. In addition, if a test for thrombophilia is positive, his or her children or siblings may want to be tested, and if testing is positive, they will need counseling. The College of American Pathologists strongly advocates that informed consent be obtained prior to testing for thrombophilia. It should also be remembered that a positive test may affect the insurability of the patient and/or his or her family members.

As a general rule, the observed clinical VTE phenotype overrides or trumps the genotype (thrombophilia test results). In other words, the clinical features of a specific case are more important than the findings of laboratory testing. For example, if a woman not on birth control pills but who has a very strong family history of thromboembolism develops an unprovoked pulmonary embolism at age 25, this patient should probably be treated for life, regardless of the results of any thrombophilia testing. Moreover, all patients with a history of VTE, including those with and without a thrombophilic

condition, require aggressive antithrombotic prophylaxis if major surgery is planned.

CONCLUSION

Testing for a thrombophilic condition should not be undertaken without full consideration of all of the consequences. Although many experts in the field of hemostasis are excited about thrombophilia testing and recommend that all patients with an unprovoked VTE undergo thrombophilia testing, other experts caution against testing because it is expensive, may be inaccurate, may be misinterpreted, and in almost all situations does not affect the long-term management of the patient. An argument can be made to test patients for AT deficiency, particularly if there is evidence of heparin resistance at the beginning of treatment. Testing for the presence of acquired APS is also recommended as these patients need repeat testing and life-long treatment appears necessary if they meet the rigorous criteria for diagnosis.

DISCLOSURES

Potential conflict of interest: Robert Gosselin participates in the Speakers Bureau and Siemens healthcare diagnostics.

REFERENCES

1. Bick RL. Hereditary and acquired thrombophilic disorders. *Clin Appl Thromb Hemost* 2006; 12: 125–135.
2. Greer IA. Inherited thrombophilia and venous thromboembolism. *Best Pract Res* 2003; 17: 413–425.
3. McKenzie SB, Clare CN, Smith LA, Sang JE. Laboratory test utilization in the diagnosis of hypercoagulability. *Clin Lab Sci* 2000; 13: 215–221.
4. Egeberg O. Inherited antithrombin deficiency causing thrombophilia. *Thromb Diath Haemorrh* 1965; 13: 516–530.
5. Nachman RL, Silverstein R. Hypercoagulable states. *Ann Intern Med* 1993; 119: 819–827.
6. Garcia de Frutos P, Fuentes-Prior P, Hurtado B, Sala N. Molecular basis of protein S deficiency. *Thromb Haemost* 2007; 98: 543–556.
7. Gandrille S, Borgel D, Sala N, et al. Protein S deficiency: a database of mutations—summary of the first update. *Thromb Haemost* 2000; 84: 918.
8. Michiels JJ, Hamulyak K. Laboratory diagnosis of hereditary thrombophilia. *Semin Thromb Hemost* 1998; 24: 309–320.
9. Hirsh J, Piovella F, Pini M. Congenital antithrombin III deficiency. Incidence and clinical features. *Am J Med* 1989; 87: 34S–38S.

10. Haemostasis and Thrombosis Task Force, British Committee for Standards in Haematology. Investigation and management of heritable thrombophilia. *Br J Haematol* 2001; 114: 512–528.
11. Reitsma PH, Rosendaal FR. Past and future of genetic research in thrombosis. *J Thromb Haemost* 2007; 5(suppl 1): 264–269.
12. Deitcher SR, Gomes MP. Hypercoagulable state testing and malignancy screening following venous thromboembolic events. *Vasc Med* 2003; 8: 33–46.
13. Marciniak E, Gockerman JP. Heparin-induced decrease in circulating antithrombin-III. *Lancet* 1977; 2: 581–584.
14. Iba T, Kidokoro A. High-dose antithrombin therapy for sepsis. *J Trauma* 2000; 49: 979–980.
15. Marder V, Matei DE. Hereditary and acquired thrombophilic syndromes. In: Colman RW, Hirsh J, Marder VJ, eds. *Hemostasis and Thrombosis*. 4th ed. Philadelphia, Pennsylvania: Lippincott Williams & Wilkins; 2001; 1245–1249.
16. Miletich J, Sherman L, Broze G Jr. Absence of thrombosis in subjects with heterozygous protein C deficiency. *N Engl J Med* 1987; 317: 991–996.
17. Koeleman BP, Reitsma PH, Allaart CF, Bertina RM. Activated protein C resistance as an additional risk factor for thrombosis in protein C-deficient families. *Blood* 1994; 84: 1031–1035.
18. van der Meer FJ, Koster T, Vandenbroucke JP, et al. The Leiden Thrombophilia Study (LETS). *Thromb Haemost* 1997; 78: 631–635.
19. Gosselin RC, Janatpour K, Larkin EC, et al. Comparison of samples obtained from 3.2% sodium citrate glass and two 3.2% sodium citrate plastic blood collection tubes used in coagulation testing. *Am J Clin Pathol* 2004; 122: 843–848.
20. Nicolaides AN, Breddin HK, Carpenter P, et al. Thrombophilia and venous thromboembolism. International consensus statement. Guidelines according to scientific evidence. *Int Angiol* 2005; 24: 1–26.
21. Juul K, Tybjaerg A, Schnohr P, Nordestgaard BG. Factor V Leiden and the risk for venous thromboembolism in the adult Danish population. *Ann Intern Med* 2004; 140: 330–337.
22. Wu O, Robertson L, Twaddle S, et al. Screening for thrombophilia in high-risk situations: a meta-analysis and cost-effectiveness analysis. *Br J Haematol* 2005; 131: 80–90.
23. Chinthammitr Y, Vos HL, Rosendaal FR, Doggen CJ. The association of prothrombin A19911G polymorphism with plasma prothrombin activity and venous thrombosis: results of the MEGA study, a large population-based case-control study. *J Thromb Haemost* 2006; 4: 2587–2592.
24. Miyakis S, Lockshin MD, Atsumi T, et al. International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). *J Thromb Haemost* 2006; 4: 295–306.
25. Dunn AS, Kaboli P, Halfdanarson T, et al. Do patients followed in anticoagulation clinics for antiphospholipid syndrome meet criteria for the disorder? *Thromb Haemost* 2005; 94: 548–554.
26. Crowther MA, Ginsberg JS, Julian J, et al. A comparison of two intensities of warfarin for the prevention of recurrent thrombosis in patients with the antiphospholipid antibody syndrome. *N Engl J Med* 2003; 349: 1133–1138.
27. Crowther MA, Johnston M, Weitz J, Ginsberg JS. Free protein S deficiency may be found in patients with antiphospholipid antibodies who do not have systemic lupus erythematosus. *Thromb Haemost* 1996; 76: 689–691.
28. ACL-A Antiphospholipid Antibody participant summary. College of American Pathologists. Northfield, IL, 2007.
29. Baca V, Montiel G, Meillon L, et al. Diagnosis of lupus anticoagulant in the lupus anticoagulant-hypoprothrombinemia syndrome: report of two cases and review of the literature. *Am J Hematol* 2002; 71: 200–207.
30. den Heijer M, Koster T, Blom HJ, et al. Hyperhomocysteinemia as a risk factor for deep-vein thrombosis. *N Engl J Med* 1996; 334: 759–762.
31. Mandel H, Brenner B, Berant M, et al. Coexistence of hereditary homocystinuria and factor V Leiden—effect on thrombosis. *N Engl J Med* 1996; 334: 763–768.
32. O'Donnell J, Mumford AD, Manning RA, Laffan M. Elevation of FVIII: C in venous thromboembolism is persistent and independent of the acute phase response. *Thromb Haemost* 2000; 83: 10–13.
33. Ray JG, Kearon C, Yi Q, et al. Homocysteine-lowering therapy and risk for venous thromboembolism: a randomized trial. *Ann Intern Med* 2007; 146: 761–767.
34. De Stefano V, Casorelli I, Rossi E, et al. Interaction between hyperhomocysteinemia and inherited thrombophilic factors in venous thromboembolism. *Semin Thromb Hemost* 2000; 26: 305–311.
35. Franco RF, Reitsma PH. Genetic risk factors of venous thrombosis. *Hum Genet* 2001; 109: 369–384.
36. Marz W, Nauck M, Wieland H. The molecular mechanisms of inherited thrombophilia. *Z Kardiol* 2000; 89: 575–586.
37. De Stefano V, Martinelli I, Mannucci PM, et al. The risk of recurrent deep venous thrombosis among heterozygous carriers of both factor V Leiden and the G20210A prothrombin mutation. *N Engl J Med* 1999; 341: 801–806.
38. Emmerich J, Rosendaal FR, Cattaneo M, et al. Combined effect of factor V Leiden and prothrombin 20210A on the risk of venous thromboembolism—pooled analysis of 8 case-control studies including 2310 cases and 3204 controls. Study Group for Pooled-Analysis in Venous Thromboembolism. *Thromb Haemost* 2001; 86: 809–816.
39. Margaglione M, D'Andrea G, Colaizzo D, et al. Coexistence of factor V Leiden and factor II A20210 mutations and recurrent venous thromboembolism. *Thromb Haemost* 1999; 82: 1583–1587.
40. Caprini JA, Goldshteyn S, Glase CJ, Hathaway K. Thrombophilia testing in patients with venous thrombosis. *Eur J Vasc Endovasc Surg* 2005; 30: 550–555.
41. Wu O, Robertson L, Twaddle S, et al. Screening for thrombophilia in high-risk situations: systematic review and cost-effectiveness analysis. Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) Study. *Health Technol Assess* 2006; 10: 1–110.
42. Heit JA, Mohr DN, Silverstein MD, et al. Predictors of recurrence after deep vein thrombosis and pulmonary

- embolism: a population-based cohort study. *Arch Intern Med* 2000; 160: 761–768.
43. Makris M, Rosendaal FR, Preston FE. Familial thrombophilia: genetic risk factors and management. *J Intern Med* 1997; 740: 9–15.
44. Mateo J, Oliver A, Borrell M, et al. Laboratory evaluation and clinical characteristics of 2,132 consecutive unselected patients with venous thromboembolism—results of the Spanish Multicentric Study on Thrombophilia (EMET-Study). *Thromb Haemost* 1997; 77: 444–451.
45. Christiansen SC, Cannegieter SC, Koster T, et al. Thrombophilia, clinical factors, and recurrent venous thrombotic events. *JAMA* 2005; 293: 2352–2361.
46. Kearon C. Long-term management of patients after venous thromboembolism. *Circulation* 2004; 110(suppl 1): I10–I18.
47. White RH, Dager WE, Zhou H, Murin S. Racial and gender differences in the incidence of recurrent venous thromboembolism. *Thromb Haemost* 2006; 96: 267–273.
48. Kyrle PA, Minar E, Bialonczyk C, et al. The risk of recurrent venous thromboembolism in men and women. *N Engl J Med* 2004; 350: 2558–2563.
49. Baglin T, Luddington R, Brown K, Baglin C. Incidence of recurrent venous thromboembolism in relation to clinical and thrombophilic risk factors: prospective cohort study. *Lancet* 2003; 362: 523–526.
50. De Stefano V, Martinelli I, Mannucci PM, et al. The risk of recurrent venous thromboembolism among heterozygous carriers of the G20210A prothrombin gene mutation. *Br J Haematol* 2001; 113: 630–635.
51. Eichinger S, Pabinger I, Stumpflen A, et al. The risk of recurrent venous thromboembolism in patients with and without factor V Leiden. *Thromb Haemost* 1997; 77: 624–628.
52. Meinardi JR, Middeldorp S, de Kam PJ, et al. The incidence of recurrent venous thromboembolism in carriers of factor V Leiden is related to concomitant thrombophilic disorders. *Br J Haematol* 2002; 116: 625–631.
53. Palareti G, Cosmi B. Predicting the risk of recurrence of venous thromboembolism. *Curr Opin Hematol* 2004; 11: 192–197.
54. Prandoni P. Risk factors of recurrent venous thromboembolism: the role of residual vein thrombosis. *Pathophysiol Haemost Thromb* 2004; 33: 351–353.
55. Simioni P, Prandoni P, Lensing AW, et al. The risk of recurrent venous thromboembolism in patients with an Arg506→Gln mutation in the gene for factor V (factor V Leiden). *N Engl J Med* 1997; 336: 399–403.
56. van den Belt AG, Sanson BJ, Simioni P, et al. Recurrence of venous thromboembolism in patients with familial thrombophilia. *Arch Intern Med* 1997; 157: 2227–2232.
57. Vossen CY, Walker ID, Svensson P, et al. Recurrence rate after a first venous thrombosis in patients with familial thrombophilia. *Arterioscler Thromb Vasc Biol* 2005; 25: 1992–1997.
58. Rodger MA, Kahn SR, Wells PS, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. *CMAJ* 2008; 179: 417–426.
59. Bauer KA, Rosendaal FR, Heit JA. Hypercoagulability: too many tests, too much conflicting data. *Hematology Am Soc Hematol Educ Program* 2002; 353–368.
60. Van Cott EM, Laposata M, Prins MH. Laboratory evaluation of hypercoagulability with venous or arterial thrombosis. *Arch Pathol Lab Med* 2002; 126: 1281–1295.
61. Bauer KA. The thrombophilias: well-defined risk factors with uncertain therapeutic implications. *Ann Intern Med* 2001; 135: 367–373.
62. Crowther MA. Antiphospholipid antibody syndrome: recent findings on managing this challenging condition. *Cleve Clin J Med* 2005; 72: 245–249.
63. Kearon C, Kahn SR, Agnelli G, et al. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). *Chest* 2008; 133(O:suppl): 454S–545S.
64. Kearon C, Julian JA, Kovacs MJ, et al., for the ELATE Investigators. Influence of thrombophilia on risk of recurrent venous thromboembolism while on warfarin: results from a randomized trial. *Blood* 2008; 112: 4432–4436.
65. Dalen JE. Should patients with venous thromboembolism be screened for thrombophilia? *Am J Med* 2008; 121: 458–463.



Giant
Resources...
Purebred
Content

Wiley InterScience® Reference Works

Sample Title List

Encyclopedia of Life Sciences

Encyclopedia of Statistical Sciences

Kirk-Othmer Encyclopedia of Chemical Technology

Sax's Dangerous Properties of Industrial Materials

Ullmann's Encyclopedia of Industrial Chemistry

Van Nostrand's Encyclopedia of Chemistry

Capabilities

Full text HTML and PDF (for some)

Reference linking

Citation download

Access

Single article or full title

One-time fee, subscription,

or Pay-Per-View options

Unlimited concurrent usage

Today's sophisticated researcher demands fast and convenient online access to the giants of scientific reference.

From Kirk-Othmer to the *Encyclopedia of Life Sciences*,

Wiley InterScience® Reference Works deliver the biggest names in scientific research literature to your desktop, right when you need them. Encompassing subjects ranging from chemistry, life sciences, and medicine to psychology, engineering, and physics, these multi-volume encyclopedias, handbooks, and dictionaries are enhanced by powerful functionality that accelerates your path to discovery. With **Wiley InterScience**, you can be sure to **Discover Something Great.**



WILEY
InterScience®
DISCOVER SOMETHING GREAT



1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 75, NO. 4
JULY/AUGUST 2009

Issue Theme
Innovations in
Medical Education

Theme Editor
Lisa Coplit, MD

A Journal of Translational and Personalized Medicine

Teaching Teamwork in
Medical Education

Simulation in Medical
Education: What's the
Evidence of Its Utility?

Student-run Health Clinics:
A Novel Arena for Teaching
Systems-based Practice

Are Medical Student Notes in
Medical Records a Liability?

Impact of the Humanities in
Medical Education

Direct Observation in
Medical Education:
Evidence for Validity

Mission-based Budgeting for
Medical Education: Is It
Ready for Prime Time?

Introducing INSPIRE, a
Scholarly Component in
Undergraduate Medical
Education



Special Features

Combined Effect of Smoking and Alcohol Use on Risk of Upper Aero Digestive Tract Cancer

Evidence-based Medicine: Accuracy of CT Angiography in Coronary Artery Disease

Management of Stage D Heart Failure

WILEY-BLACKWELL

**Newly
Relaunched!**



**Keep Abreast of the Latest Trends in
Translational and Personalized Medicine...
Enter Your Subscription Today!**

MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Editor-in-Chief:

Penny A. Asbell, MD, FACS, MBA

Mount Sinai School of Medicine, New York, NY

2008 Themes and Guest Editors

Quality of Healthcare and Racial
Health Disparities (Mark Chassin)

9/11 World Trade Center Disaster:
Past and Future (Robin Herbert,
Jacqueline Moline)

Anxiety and Depression:
Leading Edge of Therapy
(Dennis Charney, Mathew Sanjay)

Diabetes Type 1:
New Light on Etiology,
New Approaches to Management
(Derek Leroith)

Hospital Medicine:
An Ascending New Field
(Andrew Dunn)

Advanced Lung Disease:
Meeting the Challenges
(Maria Padilla)

Continuously published since 1934 and included in *Index Medicus* from the very first volume, the **Mount Sinai Journal of Medicine (MSJM)** has a distinguished career as a general medical journal for a worldwide audience.

Building upon this tradition of education across medical specialties, **MSJM** was re-launched in 2007 as a journal of translational and personalized medicine, two disciplines in the forefront of modern healthcare. **MSJM** features articles on the evolving nature of clinical care, its close linkage to bio-medical research and the resulting opportunities for enhanced patient care.

Dedicated to improving communication among all participants in medicine today, **MSJM** is an excellent source of up-to-date medical information for healthcare clinicians, educators, researchers and policy makers worldwide. **MSJM** will inspire the creative thinking needed to create new ways to understand, prevent, and treat human disease.

Each issue includes several quality, in-depth review articles related to a single theme and several others on subjects of diverse clinical interest. All articles are written by renowned authorities in their fields, from both Mount Sinai and elsewhere, under the supervision of Guest Editors.

Enter Your 2009 Subscription Today!



Your 2009 subscription will include the six 2009 print issues of the **Mount Sinai Journal of Medicine** (Volume 76), plus concurrent full-text online access to these six issues as they are posted on Wiley's Internet publishing platform.

Two Ways for Individuals to Subscribe:

1. Mount Sinai Alumna/Alumnus:
\$40.00 (Worldwide)
2. Personal Subscription:
\$95.00 (U.S./Canada/Mexico)
\$120.00 (Outside North America)

Complimentary Online Access for Institutions!

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

PERSONAL SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 personal subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$95.00 (U.S./Canada/Mexico) ☐ \$120.00 (Outside North America)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

Method of Payment (Canada residents please add GST. Personal rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **John Wiley & Sons, Inc.** Payment must be made in U.S. dollars and drawn on a U.S. bank.)
☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA

Card # _____ Exp. Date _____

Signature _____
(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM0909

MSJM MEMBER SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 Alumna/Alumnus Subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$40.00 (Worldwide)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

METHOD OF PAYMENT (Alumna/Alumnus rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **The Mount Sinai School of Medicine**. Payment must be made in U.S. dollars and drawn on a U.S. bank.)
☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA ☐ Other

Card # _____ Exp. Date _____

Signature _____
(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM09

COMPLIMENTARY ONLINE ACCESS FOR INSTITUTIONS

Mount Sinai Journal of Medicine

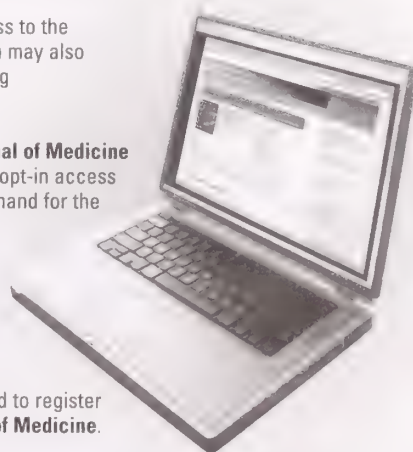
A Journal of Translational and Personalized Medicine

2009. Volume 76. 6 issues per year. Print ISSN: 0027-2507. Online ISSN: 1931-7581.

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

Institutional online and print pricing for the **Mount Sinai Journal of Medicine** will be set in 2009, subsequent to introductory complimentary opt-in access during 2008 and 2009. Pricing will be based on the overall demand for the Journal and on the size and type of the subscribing institution (e.g. academic, government, hospital, or corporate). Institutions that receive complimentary online access during 2008 and 2009 will have archival access in subsequent years according to Wiley's licensing terms and conditions.

Ask your librarian to e-mail optinaccess@wiley.co.uk or visit www.interscience.wiley.com/msjregister for more details and to register for complimentary online access to the **Mount Sinai Journal of Medicine**.





NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS MAIL

PERMIT NO. 2277

HOBOKEN, NJ

POSTAGE WILL BE PAID BY ADDRESSEE

JOHN WILEY & SONS INC
ATTN: PETER JANSIC, MAIL STOP 8-02
111 RIVER STREET
HOBOKEN, NJ 07030-9441



PLACE
POSTAGE
STAMP
HERE

INES SANTOS
MOUNT SINAI JOURNAL OF MEDICINE
ONE GUSTAV L. LEVY PLACE BOX 1094
NEW YORK, NY 10029-6574

**Library
Routing
Card**

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editors

Dennis S. Charney
Maria New
C. Warren Olanow

Associate Editors

Stuart Aaronson
David H. Adams
Erwin Bottinger
Eric Hollander
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Fred Lublin
Michael L. Marin
Thomas G. McGinn
Ira Nash
Rosamond Rhodes
Hugh Sampson
Mary Sano
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Guest Editor

Lisa Coplit, MD
Director, Institute for Medical Education
Department of Medical Education
Mount Sinai School of Medicine

Editorial Director

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com

Innovations in Medical Education



Themes for future issues of *MSJM*, each shown with Theme Editor(s):

- Zoonotic infections (Stephanie Factor)
- Obstetrics/gynecology (Rhoda Sperling)
- Alzheimer's disease (Nikos Robakis)
- Drug discovery in academic settings (Ravi Iyengar)
- Parkinson's disease (Michele Tagliati, Ronald Alterman)
- Endovascular surgery (Michael Marin)
- Genes, race, ethnicity and the environment (Gary Butts)

©2009 Mount Sinai School of Medicine. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: \$95.00 in North America, \$120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS

VOL 76, NO 4

Theme Introduction

- 315** Innovations in Medical Education: Getting Closer to Our Goals

Lisa Coplit

Theme Articles

- 318** Teaching Teamwork in Medical Education

Susan Lerner, Diane Magrane, Erica Friedman

- 330** The Utility of Simulation in Medical Education: What Is the Evidence?

Yasuharu Okuda, Ethan O. Bryson, Samuel DeMaria Jr., Lisa Jacobson, Joshua Quinones, Bing Shen, Adam I. Levine

- 344** Student-run Health Clinics: A Novel Arena to Educate Medical Students on Systems-based Practice

Yasmin S. Meah, Eric Smith, David C. Thomas

- 357** Medical Student Documentation in the Medical Record: Is It a Liability?

Peter Gliatto, Philip Masters, Reena Karani

- 365** Direct Observation in Medical Education: A Review of the Literature and Evidence for Validity

H. Barrett Fromme, Reena Karani, Kathel Dunn, Steven M Downing

- 372** Evaluating the Impact of the Humanities in Medical Education

Edward J. Ronan, Mary R. Rifkin, Jeremy S. Abramson, Andrea Wershof Schwartz, Israel Wojnowich, Robert Accordino

- 381** Mission-based Budgeting for Education: Ready for Prime Time?

Michelle Sainte, Steven L. Kanter, David Muller

- 387** Introducing *INSPIRE*, a scholarly component in Undergraduate Medical education

Karen Zier, Lisa D. Coplit

Special Features

- 392** The Risk of Upper Aero Digestive Tract Cancer associated with Smoking, with and without Concurrent Alcohol Consumption

Alireza Ansary Moghaddam, Rachel R. Huxley, TH Lam, Mark Woodward

- 404** The Management of Stage D Heart Failure

Mark Harrison, Anelechi Anyanwu, Sean P. Pinney

- 415** Evidence-based Medicine: The Accuracy of CT Angiography for Coronary Artery Disease

Deborah Korenstein, Thomas McGinn



HINARI PHASE I COUNTRIES

Free access to more than 100 biomedical journals published by Wiley



HINARI PHASE II COUNTRIES

Nominal cost access to more than 100 biomedical journals published by Wiley

Developing World Access to Research

“perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries.”

Former WHO Director-General Dr Gro Harlem Brundtland (July, 2001)

Wiley, together with the World Health Organisation and fellow Scientific, Technical and Medical publishers, is proud to play a role in furthering global dissemination of biomedical information.

Through the Health InterNetwork Access to Research Initiative (HINARI) www.healthinternetwork.org, free or nominal cost online access to more than 2,200 biomedical journals from 42 publishers is provided to more than 1,000 institutions in approximately 100 of the world's lowest income countries.

Wiley is one of the six founding publishers who led this initiative with the WHO. We offer more than 100 of our leading biomedical journals through the HINARI program; including many leading journals published with our society partners.

HINARI
ACCESS TO RESEARCH



WILEY
InterScience®
DISCOVER SOMETHING GREAT
www.interscience.wiley.com

Wiley – helping to make a world of difference.

THEME INTRODUCTION

Innovations in Medical Education: Getting Closer to Our Goals

Lisa Coplit, MD

Department of Medical Education, Mount Sinai School of Medicine, New York, NY

For nearly a century, despite revolutionary advances in science, the goal of medical education has never changed: to provide a learning environment that produces highly competent, compassionate, and altruistic physicians who are prepared to care for patients and use scientific methods to approach disease prevention, treatment, and cure. However, the means to that end must evolve because the requisite competencies, the context in which students learn, and the expectations for how they practice have dramatically shifted. In 1910, the Flexner Report from the Carnegie Foundation for the Advancement of Teaching fueled the first major reforms in medical education in the United States.¹ Medical schools were held to higher academic standards, the biomedical model of teaching was established, and academic medical centers became the ideal clinical training venue for quality teaching, research, and patient care. Although these changes were critically important at the time, Cooke *et al.*² proposed that subsequent changes in medicine, such as the pressure on both researchers and clinicians to increase their productivity, have depleted our pool of available teachers who can offer outstanding clinical teaching and role-model the application of research to patient care. Many other factors have influenced medical education, including the exponential increase in our understanding of disease, diagnosis, and treatment, which requires physicians to shift from being

autonomous practitioners to being effective partners of interprofessional teams. In addition, the increasing accountability to the public for the quality of care provided by physicians, students' financial indebtedness, and the conflicts of interest that they experience, which may erode their professional values, have also had great impacts. The magnitude and power of these factors on the public and professional expectations of physicians have brought us to another pivotal juncture because, as Cooke *et al.* explained, "our approach to education is inadequate to meet the needs of medicine." They conclude that medical schools must ensure that teaching is contextual, learning facilitates students' analysis, integration, and application of research/literature to patients, professionalism is explicitly incorporated into the curriculum, and rigorous assessments of competency in all domains are developed.

Darrel G. Kirch, MD, in his 2007 address to the American Association of Medical Colleges,³ added that the current individualistic culture of medicine must adapt to the needs of our patients who require coordination of their care. Our current "discontinuous assortment of independent lectures and tests" must give way to an "integrated learning continuum" if we are to train truly competent physicians. He also stated that we must maintain the scientific foundation of medicine but will need to embrace these other values to truly meet our goals.

So how are we adapting to these necessary changes? Instead of a 1-size-fits-all model, various innovations in medical education are emerging around the country to address the needs of our students and patients. Educators are increasingly approaching teaching and learning as Flexner would have likely suggested. They are basing their decisions on the literature, creating and sharing interventions that are novel, assessing the effects on learners, and disseminating their work for peer review. Sharing our innovations with others defines scholarship in

Address Correspondence to:

Lisa Coplit, MD

Department of
Medical Education

Mount Sinai School of Medicine
New York, NY

Email: lisa.coplit@mssm.edu

education³ and will hopefully influence the reforms being created at other institutions and nationally. In this medical education theme issue of the *Mount Sinai Journal of Medicine*, the authors provide us with comprehensive reviews of important issues and concepts that address the national call for change in medical education at the undergraduate, graduate, and continuing medical education levels of training.

Lerner *et al.* detail what is known about teaching teamwork in undergraduate medical education. They preface the utility of teamwork training for medical students with information from other industries, such as aviation and business, and then explain how students are being taught these skills and the methods being used to assess their effectiveness as team members. Okuda *et al.* provide us with an in-depth review of the evidence for using simulation in medical education. Simulation addresses the issue of the accountability of the education system for the safety of patients while providing a high-quality means of assessing learners' skills in many domains, including medical knowledge, procedural skill, professionalism, and teamwork. Fromme *et al.* give us a detailed analysis of the validity of direct observation as a tool for evaluating the clinical competency of medical students and residents. Direct observation has the benefit of assessing learners at the highest level when they are applying their knowledge, skills, and attitudes in real patient care settings. Meah *et al.* describe the emergence of student-run health clinics, which typically serve underserved and underinsured populations of patients. These clinics may take us 1 step further toward offering the clinical context necessary to teach students how to care for patients within our complex healthcare delivery system while role-modeling the values of collaboration and patient advocacy. Gliatto *et al.* examine the growing trend toward limiting medical student documentation in patients' medical records. This article is unique in that it discusses a firmly established educational intervention that is in jeopardy of being eliminated but that may be of particular value as we seek to teach students about communication skills and accountability and may afford clinical teachers a valuable method of assessing these skills. Schwartz *et al.* review the literature on the benefits of including humanities in medical education. In an effort to offer students explicit teaching in desired attitudes of the medical profession (empathy, professionalism, patient centeredness, and altruism), many institutions are either seeking potential students with a background in the humanities or infusing humanities into their curricula. We look forward to data about the effects of these

curricula on helping maintain the ideals of our profession. Lastly, Sainte *et al.* review the history, current practices, and potential for mission-based budgeting, a system that attempts to accurately align costs with faculty activities. They focus on the complexities and advantages of creating a metric to support the educational component of this payment structure. Adequately and accurately acknowledging faculty contributions to the teaching missions of our institutions is another important step toward retaining outstanding educators, which is crucial for maintaining excellence and for initiating improvements in our educational system.

In a special article that serves as a spotlight for an educational innovation at Mount Sinai School of Medicine, Zier and Coplit describe the Individual Scholarly Project and Independent Research Experience (INSPIRE), the first implementation of Mount Sinai School of Medicine's curriculum reform. INSPIRE allows fourth-year students to conduct an independent scholarly project with the goal of producing students who are self-directed, lifelong learners. Scholarly programs are another example of a curricular innovation that is rapidly being embraced at medical schools across the country in order to produce the next generation of physician-scholars and leaders in medicine. Supporting our students' scholarly explorations of their personal interests in medicine will provide them with a richer educational experience that enables them to critically evaluate the scientific literature and generate new knowledge in order to deliver the highest quality clinical care to patients.

As medical education has been transformed, so have the expectations of the educators. Educational scholarship is increasingly valued, thus promoting the work and recognition of medical educators.⁴ However, just as Flexner demanded for patient care, educators are also expected to take an evidence-based approach to teaching, learning, and assessing competency. The articles in this issue provide scholarly contributions to the larger educational community and represent a small sample of curricular innovations that are being examined and developed to meet the challenges of our profession. With the creativity, dedication, and enthusiasm of so many educators at the local and national levels, we are poised for a similar explosion in educational methodologies and outcomes that will prepare the physicians of the 21st century.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Beck AH. The Flexner Report and the standardization of American medical education. *JAMA* 2004; 291: 2139–2140.
2. Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner Report. *N Engl J Med* 2006; 355: 1339–1344.
3. Kirch DG. AAMC president's address 2007: culture and the courage to change. Available at: <http://www.aamc.org/meetings/annual/2007/highlights/president.htm>. Accessed May 2009.
4. Association of American Medical Colleges. Advancing educators and education: defining the components and evidence of educational scholarship. Summary report and findings from the AAMC Group on Educational Affairs Consensus Conference on Educational Scholarship. Available at: <http://www.aamc.org/members/gea/educ-scholar.pdf>. Accessed May 2009.

Teaching Teamwork in Medical Education

Susan Lerner, MD,¹ Diane Magrane, MD,² and Erica Friedman, MD¹

¹Mount Sinai Medical Center, New York, NY

²Association of American Medical Colleges, Washington, DC

ABSTRACT

Teamwork has become a major focus in healthcare. In part, this is the result of the Institute of Medicine report entitled *To Err Is Human: Building a Safer Health System*, which details the high rate of preventable medical errors, many of which are the result of dysfunctional or nonexistent teamwork. It has been proposed that a healthcare system that supports effective teamwork can improve the quality of patient care and reduce workload issues that cause burnout among healthcare professionals. Few clear guidelines exist to help guide the implementation of all these recommendations in healthcare settings. In general, training programs designed to improve team skills are a new concept for medicine, particularly for physicians who are trained largely to be self-sufficient and individually responsible for their actions. Outside of healthcare, research has shown that teams working together in high-risk and high-intensity work environments make fewer mistakes than individuals. This evidence originates from commercial aviation, the military, firefighting, and rapid-response police activities. Commercial aviation, an industry in which mistakes can result in unacceptable loss, has been at the forefront of risk reduction through teamwork training. The importance of teamwork has been recognized by some in the healthcare industry who have begun to develop their own specialty-driven programs. The purpose of this review is to discuss the current literature on teaching about teamwork in undergraduate medical education. We describe the science of teams, analyze the work in team training that has been done in other fields, and assess what work has been done in other fields

about the importance of team training (ie, aviation, nonmedical education, and business). Additionally, it is vital to assess what work has already been done in medicine to advance the skills required for effective teamwork. Much of this work has been done in fields in which medical professionals deal with crisis situations (ie, anesthesia, trauma, and labor and delivery). We describe the current programs for teaching medical students these essential skills and what recommendations have been made about the best ways to introduce teaching this skill set into the curriculum. Finally, we include a review on assessing teamwork because one cannot teach team training without implementing an assessment to ensure that the skills are being learned. *Mt Sinai J Med* 76:318–329, 2009. © 2009 Mount Sinai School of Medicine

Key Words: crew resource management, healthcare, medical education, safety, simulation, team performance, team training, teamwork.

Teamwork has become a major focus in healthcare. The impetus for this new interest in teamwork is the Institute of Medicine (IOM) report entitled *To Err is Human: Building a Safer Health System*,¹ which details the high rate of preventable medical errors, many of which are the result of dysfunctional or nonexistent teamwork. This report suggests that teamwork is required for effective patient management because of the increased specialization of tasks, the increased complexity and risks associated with treatment options, and the need to ensure appropriate healthcare outcomes and patient safety. As the report states, “beyond their cost in human lives, preventable medical errors exact other significant tolls . . . total costs of between \$17 billion and \$29 billion per year . . . loss of trust in the healthcare system by patients and diminished satisfaction by both patients and health professionals.” A variety of factors have contributed to this tragedy, but an important component seems to be the “decentralized and fragmented nature of the healthcare delivery system—a nonsystem,” that is, a system that lacks

Address Correspondence to:

Susan Lerner

Mount Sinai Medical Center

New York, NY

Email: susan.lerner@mountsinai.

org

essential organization.¹ It has been proposed that a healthcare system that supports effective teamwork can improve the quality of patient care and reduce workload issues that cause burnout among healthcare professionals.²

One of the problems identified by the IOM was "the lack of a single designated government agency devoted to improving and monitoring safety throughout the healthcare delivery system."¹ It went on to suggest that

Congress should create a Center for Patient Safety that would set national safety goals and track progress in meeting them; develop a research agenda; define prototype safety systems; develop, disseminate, and evaluate tools for identifying and analyzing errors; develop methods for educating consumers about patient safety; and recommend additional improvements as needed.

The IOM also concluded that research has not yet developed a comprehensive model of team training performance in medical settings.¹ However, analyzing the science of team performance and training can help the medical community to improve patient safety. In addition, its report suggests implementing and regulating team training throughout each healthcare provider's career.³

In his November 2007 Association of American Medical Colleges (AAMC) presidential address, "Culture and the Courage to Change," Darrell Kirch, MD, argued that organizational culture is every bit as important as organizational strategy for academic medical centers. Kirch called for a redirection of the medical culture to one that is focused upon "collaboration, shared accountability, and team performance".⁴ Basic assumptions of how the medical community provides the highest quality care with respect to patient safety are changing, and team approaches are central to those changes.

Therefore, it is imperative to develop and implement a curriculum that teaches and promotes teamwork within medical education. Medical organizations and teaching institutions must take the lead in developing and teaching team training in healthcare, but there is insufficient direction about how this should take place. Most teamwork training in medicine has been done in fields that deal with crisis situations (ie, anesthesia, trauma, and labor and delivery). This article reviews the science of team performance and team training as these provide the foundation for creating models that will work in the healthcare profession. We analyze the efforts in other fields (ie, aviation, nursing, and business) and assess the contribution and applicability of these efforts in medicine. We conclude by describing existing

programs for teaching medical students about these essential skills and by providing recommendations for introducing the teaching of this skill set into the curriculum.

DEFINITION AND SCIENCE OF TEAMS

In interviews with hundreds of people in more than 50 different teams in 30 companies, Katzenbach and Smith discovered what differentiates various levels of team performance, where and how teams work best, and how to enhance their effectiveness, and they wrote this in their book, *The Wisdom of Teams: Creating the High-Performance Organization*.⁵ They developed the frequently quoted definition of a team as "a small number of people with complementary skills who are committed to a common purpose, performance goals, and approach for which they hold themselves mutually accountable." Moreover, they wrote that "teams outperform individuals acting alone or in larger organizational groupings, especially when performance requires multiple skills, judgments, and experiences."

Teams must also have meaningful task interdependencies (ie, the job cannot be done by a single individual), and usually task-relevant knowledge is distributed among the team members.⁶ In addition, researchers have described the knowledge, skills, and attitudes (ie, competencies) necessary for effective team performances. These team competencies allow organizations to establish the appropriate requirements for their teams and strategies to enhance teamwork and performance. Knowledge-based competencies (what team members know) refer to the necessity of understanding facts, concepts, relations, and underlying foundations of information that a team member must have to perform a task. Skill-based competencies (what team members do) are the necessary behavioral sequences and procedures needed for task performance. Attitude-based competencies (how team members feel) refer to the affective components needed to perform the task. Teamwork is greatly affected by how team members feel about one another and the task.⁶

Even though all teams are different in terms of membership, group dynamics, and goals, they go through similar processes as they develop from an immature team into one that is mature and productive.⁷ Wells,⁸ a writer about teaching teamwork in information systems, stated that teams are thought to develop in stages. These stages are team forming, team building, team management, and team/project assessment. She suggested that there are several conditions that promote team effectiveness.

Teams should be formed as early as possible to allow some time for the team building stage before the real work begins. In addition, many believe that teams should not be self-selected because this tends to produce homogeneous units. Also, Katzenbach and Smith⁵ stated that

Conflict, like trust and interdependence, is also a necessary part of becoming a real team. Seldom do we see a group of individuals forge their unique experiences, perspectives, values, and expectations into a common purpose, a set of performance goals, and approach without encountering a significant conflict. And the most challenging risks associated with conflict relate to making it constructive for the team instead of enduring it.

The effectiveness of a team's work depends on the internal group processes: the ways in which conflicts are managed, the amount of trust between members, and the use of an appropriate leadership style.⁷

Katzenbach and Smith⁹ described further in an article in the *Harvard Business Review* "what makes the difference between a team that performs and one that doesn't."⁹ They proposed that

The essence of a team is common commitment. Without it, groups perform as individuals; with it, they become a powerful unit of collective performance. This kind of commitment requires a purpose in which team members can believe. By contrast, failed teams rarely develop a common purpose.

Finally, they added that no group ever becomes a team until it can hold itself accountable as a team.

Kirkman and Rosen¹⁰ studied 100 teams in 4 organizations and described the association between team empowerment and team effectiveness. They concluded that empowered teams share 4 experiences: potency, meaningfulness, autonomy, and impact. They suggested that team members must believe in their team's capabilities, find meaning in their team's tasks, and fully realize the impact of their work if they are to become truly high-performance teams. Teams need supportive leaders who will trust them to make important decisions, celebrate their successes, and learn from their mistakes. In addition, they suggested that leaders of empowered teams must play the role of coaches and facilitators, helping teams define tasks, structure activities, and monitor their own progress.

Evaluating the effectiveness of a team requires that the team members be willing and equipped to engage in a process of team evaluation and self-analysis. Team evaluation involves assessing both the process (ie, team interactions) and outcomes (ie, accomplishments). Criteria need to be established,

and the team processes and outcomes need to be measured against these criteria to confirm and document team effectiveness.⁷ In studying the effectiveness of building designers, Busseri and Palmer¹¹ showed a positive correlation between ratings of the team's design and ratings of the team's processes. They suggested that assessing team performance during a task positively affects how well the team functions. Teams that periodically self-assess their teamwork processes will work more effectively and have higher levels of member satisfaction than teams that do not.

Teamwork can be defined as the ability of team members to work together, communicate effectively, anticipate and meet one another's demands, and inspire confidence resulting in a coordinated collective action.⁶ *Collective action* is defined as the pursuit of a goal or set of goals by more than 1 person. According to Salas and Cannon-Bowers,⁶ teamwork is a critical component of team performance and is composed of 4 key behavioral characteristics. These are performance monitoring (the ability to apply appropriate task strategies to develop common understandings of the team environment, including an understanding of teammate workload, fatigue, stress, and skills and the environment external to the team itself), feedback, closed-loop communication (the exchange of information between a sender and a receiver), and backup behaviors (a person's ability to anticipate other team members' needs through knowledge about their responsibilities).

Researchers and medical professionals agree that patient treatment and safety are improved through multidisciplinary teamwork, but as Salas *et al.*¹² stated, "the reality is that teamwork occurs infrequently, is fraught with difficulties, and is misunderstood." It does not require that one like or feel close to one's team members; however, it does rely on one's willingness to cooperate for a shared goal. In the healthcare field, that goal is maintaining the patient's health status and avoiding errors.¹² Teamwork is a set of interrelated behaviors, cognitions, and attitudes. It is more than knowledge and skills. Teamwork depends on the ability of each team member to (1) anticipate the needs of others, (2) adjust to each other's actions and to the changing environment, and (3) have a shared understanding of how a procedure should happen in order to identify when errors are occurring and how to correct for these errors.¹²

Another challenge of team management is not simply to execute existing processes efficiently; it is also imperative to implement new processes and to learn how to implement new processes

quickly. Edmondson *et al.*¹³ described in the *Harvard Business Review* how her group studied surgical teams at 16 major medical centers to see how they implemented a difficult new procedure for performing cardiac surgery. What they found was evidence about a team's ability to adapt to a new way of working. First, they identified factors that did not seem to matter in how quickly teams learned the new techniques: the educational background and surgical experience of the team, high-level management support for the new technology, the status of the surgeon, and finally debriefings and project audits. What did matter was the way in which teams were put together and how they drew on their combined experiences. Teams that learned the new procedure most quickly shared 3 essential characteristics: they were designed for learning, their leaders framed the challenge in such a way that team members were highly motivated to learn, and the leaders' behavior created an environment of psychological safety that fostered communication and innovation.

TEAMWORK TRAINING

Teamwork is not an automatic consequence of placing people together.¹² Team members need training to learn how to work together and understand the professional role and responsibility of each person.⁴ Team training can be defined as a set of tools and methods that form an instructional strategy in combination with requisite competencies and training objectives. Team training has 1 basic objective: to affect crucial team competencies.⁶ Changes in how healthcare professionals are trained to view teamwork may reduce conflict and improve team performance. Optimal interdisciplinary training begins early in medical education.¹² This concept of training about teamwork has been embraced by the medical community. The Accreditation Council for Graduate Medical Education echoes that teamwork is a skill that needs to be taught and assessed. It identifies interpersonal and communication skills as 1 of 6 core competencies in graduate medical education and requires residents to demonstrate skills that "result in effective information exchange and teaming with . . . professional associates."¹⁴ The AAMC issued the Medical School Objectives Project, which sets forth "the ability to communicate effectively with patients, patients' families, colleagues, and others with whom physicians must exchange information in carrying out their responsibilities" as one of the core skills that students should possess prior to graduation from medical school.¹⁵ Team training is also currently suggested as part of a comprehensive

patient safety plan published by the Joint Commission on Accreditation of Healthcare Organizations, the regulatory agency charged with hospital accreditation in the United States.¹⁶

Few guidelines exist to help implement these recommendations in healthcare settings. In general, programs designed to improve team skills are a relatively new concept for medicine, particularly for physicians who are trained largely to be self-sufficient and individually responsible for their actions.¹⁶ Additional training needs to be done to promote interprofessional training in medicine. There needs to be a willingness to work together toward the common goal of patient safety. If a team member decides to work independently, then the effectiveness and success of the team are placed in jeopardy.¹⁷

Outside of healthcare, research has shown that teams working together in high-risk and high-intensity work environments make fewer mistakes than individuals. This evidence arises from commercial aviation, the military, firefighting, and rapid-response police activities. Simulation is the primary educational component of many of these programs. Simulation refers to the re-creation of an actual event that has previously occurred or could potentially occur. One of the greatest values of simulation is that it can be used over and over again to perfect an action, a procedure, or a conversation without the providers or patients ever being exposed to harm.¹⁸ Debriefing is of vital importance after a team simulation experience. Debriefing allows the team to learn where errors occurred and how they could potentially have been prevented, and it also allows recognition of areas of appropriate performance. Simulation without debriefing has been shown to be ineffective because errors can be repeated if team members have not been informed that they were making mistakes.¹⁸

Commercial aviation is an industry in which mistakes can result in unacceptable loss. This industry has been at the forefront of risk reduction through teamwork training. The aviation industry and the National Aeronautics and Space Administration determined that the human error aspects of the majority of air crashes were due to failures of interpersonal communication, decision making, and leadership. As a result, training programs were developed to make communication more effective on the flight deck in order to reduce error.¹⁹ One of the best known team training strategies to emerge from the aviation setting is crew resource management (CRM) training. This training focuses on behavioral skills and teamwork training.²⁰ CRM teaches that all members of a team are vital. If a team member at any level believes that something is not being done

appropriately or in the best interest of the team or other people that have put their trust in the team, then that member must speak up. The fact that humans make mistakes but are able to learn from these errors and prevent their repetition is an important tenet of CRM training.²⁰

Although some of the concepts of CRM may be communicated in a didactic setting, experts in CRM training suggest that they should be practiced simultaneously with more technical skills, often with high-fidelity simulators. The cognitive stress associated with crisis situations can be created or re-created through the use of simulated events to determine where, when, and why errors occur in these situations.²⁰ Pilots are given feedback not only on their technical skills and decision making strategies but also on crew management and resource utilization.¹⁸ CRM programs teach aviation crews about human limitations and involve participants in assessments of their own behavior and the behavior of their peers. Concepts that are taught include those related to making inquiries, seeking relevant data, advocating actions, communicating proposed action plans, resolving conflicts, and making decisions. Additionally, participants learn to understand cognitive errors and are taught how stressors, such as fatigue, crisis, and work overload, contribute to human error.¹⁹

Many healthcare organizations have set up team training programs based on the principles of CRM.²⁰ In particular, CRM training programs have been used in operating rooms, anesthesia departments, emergency departments (EDs), intensive care units, and labor and delivery units. The principles of CRM and their effect on safety have not been completely evaluated in aviation or in healthcare.¹⁹ Researchers caution that although many of the lessons learned in aviation will have counterparts in medicine, trying to directly apply one industry's practices to the other risks both the application of unwarranted elements and the omission of aspects specific to medical practice.¹⁸

Military service is a second high-risk context in which the consequences of error can be dire. Military team training research is fairly recent and led to the development of a program known as Tactical Decision Making Under Stress (TADMUS) in 1990. The program was launched as a result of 2 tragic incidents at sea. Specifically, a missile struck the USS Stark in 1987, resulting in the deaths of 37 American servicemen. This incident led to a heightened readiness and tension aboard all US military ships operating in the Persian Gulf. In July 1988, with tensions in the area already high, the USS Vincennes, mistakenly believing it was under

attack, shot down an Iranian airbus, killing all 290 people onboard. The Navy investigation concluded the commanding officer had acted properly, given all the information he had received, but noted that stress, task fixation, and unconscious distortion of data may have played a significant role in the incident. The investigations into these 2 incidents revealed the need to know more about the behavior of teams and to improve decision making under stress. The TADMUS project was launched, and Navy researchers began the comprehensive task of studying teams to determine "what makes a team of experts an expert team."²⁰ Through observations of highly functioning teams and poorly functioning teams, the researchers were able to identify specific dimensions of effective teamwork.

Team dimensional training (TDT) incorporates these research findings and focuses on 4 areas (or dimensions) essential for effective teamwork. These areas are information exchange, communication, supporting behavior, and initiative/leadership. TDT helps team members to diagnose and correct their own performance problems, thereby enabling them to adapt quickly to unfolding events and to learn from and build upon their previous experiences together. The TDT process involves a structured approach, which includes a prebriefing observation of team performance, a diagnosis of performance, and a debriefing. In the prebriefing, the team members are introduced to or reminded of the various team dimension elements. Observations of team performance may be made during training exercises or actual incidents. The observations are recorded for use in the debriefing session. Generally, a tactical debriefing is conducted first to clarify and familiarize all team members with the events that took place and to address tactical areas for positive reinforcement and/or improvement. The TDT debriefing does not address tactical issues but focuses the team on the various dimensions of effective teamwork. This process involves self-analysis by the team and is facilitated in a way that allows the team to self-correct problem areas. Goals are set by the team to work on improving problem areas of teamwork. The process helps teams to analyze their operational mistakes while teaching team leaders to guide their members through the self-correction process. New training development and delivery technology, including scenario generation software, virtual environments, and distributed simulation facilities, has made it possible for widely dispersed personnel and units to train together and to exchange feedback.²⁰

The oil and gas industry is also striving to improve health and safety performance to reduce errors and incidents. Oil and gas exploration

uses drilling project teams who come together to develop a planning program for drilling a well and overseeing drilling operations. Crichton²¹ studied drilling team members and identified several of the same themes that have arisen elsewhere in the context of a team working in a complex, uncertain, hazardous environment. He identified the specific areas (teamwork, leadership, stress, and communication) that could benefit from training interventions in order to maximize safe and effective performance.

TEAMS IN HEALTHCARE

In healthcare, common examples of teams include project teams (eg, quality improvement teams), care delivery teams, and management teams. Care delivery teams can be subdivided according to patient population (eg, geriatric teams), disease types (eg, stroke teams), or care delivery settings (operating room or intensive care unit).² A wide range of collaboration can occur, depending on the type of care required. Higher levels of collaboration are often needed when complex medical problems require care by several healthcare professionals. Collaboration among health professionals forms a dynamic spectrum ranging from independent parallel practice to consultation and referral (where health professionals exchange information) to interdependent coprovision of care with interdependent decision making.² The multidisciplinary team allows for each discipline to independently contribute its particular expertise to an individual patient's care. Team members in this situation work in parallel to one another, and direct communication between disciplines is minimal except through the physician in charge. Members of the team may be of the same or different disciplines. The interdisciplinary team refers to a team whose members work together closely and communicate freely to optimize care for the patient. The team is organized around solving a common set of problems instead of being organized around a single physician. Each member of the team contributes his knowledge and skill set to augment and support the others' contributions.²

CRM principles borrowed from aviation have been applied to the development of programs, both simulator-based and non-simulator-based, to train medical teams. High-acuity environments, such as the operating room, the intensive care unit, the emergency room, and the labor and delivery suites, are constantly changing, complex environments in which it is crucial that all personnel operate as a team to maintain patient well-being and safely deliver

healthcare. A subset of CRM concepts, when applied to high-acuity situations in healthcare, is called crisis resource management.¹⁷ In order for these programs to be successful, organizations must acknowledge that errors occur and deal with errors in a nonpunitive manner in which errors are viewed as an opportunity for learning and improvement.¹⁹

The Anesthesia Crisis Resource Management (ACRM) program, developed by Gaba and colleagues,²² is one of the most well-known simulator-based team training programs. In 1987, Gaba and his team began a set of studies of decision making by anesthesiologists, using a patient simulator that they invented. In those experiments, anesthesiologists of different levels of experience managed a simulated patient during a surgical procedure in which multiple medical and equipment faults were triggered. An analysis of videotapes from these early experiments suggested that the training of anesthesiologists contained gaps concerning several aspects of decision making and crisis management that were not systematically taught during training. In 1989, they began to develop a simulation-based curriculum modeled in part on the key principles of CRM in aviation. The first course using the new curriculum was held in September 1990 and was called ACRM. They considered each discipline in healthcare to be a crew containing 1 or more individuals. Several crews may work together closely as a team. For example, the operating room team consists of an anesthesia crew, a surgery crew, and a nursing crew. In ACRM, they opted for a strategy of training crews to work in teams. ACRM uses crew training to provide comprehensive teaching and practice in the integrated use of technical, cognitive, and behavioral skills in managing crises relevant to their domain. Training crews to work in teams also provides a degree of cross-discipline understanding by allowing participants to discuss other team members' views of the same activity or event. In this version of the CRM, simulated scenarios are planned to evaluate teamwork and team performance with planned feedback sessions for process evaluation. A number of behaviors are evaluated, including how efficiently the team performs, how the teams build trust among members, and how members work toward the common goal of patient safety. The debriefing session allows team members to recognize and reinforce appropriate performance while understanding how mistakes occurred and collaborating on how to prevent these errors from recurring in the future. The debriefings last approximately 40 minutes and allow participants to explore alternatives and to recognize and discuss principles of ACRM as they were either executed or ignored in the scenario.

The operating room is a high-risk environment requiring close coordination among all team members. Important characteristics of a reliable team include the ability to adapt to changes within the work environment, to maintain open and flexible communication, and to anticipate the needs of each member of the team. The operating room is a diverse environment, with staff and trainees from different disciplines. Although each of these team members is well trained in clinical task work, most are not explicitly trained in applying these skills in a team environment in which there is a high degree of task interdependence and team coordination. CRM was used in Kantonsspital at the University of Basel (Basel, Switzerland) to develop the Team Oriented Medical Simulation program. This model includes a variety of operating room personnel participating in the simulation simultaneously.¹⁷

Simulation-based operating room team training was further studied by Paige *et al.*²³ A simulation-based interdisciplinary team training program for the operating room was developed by the authors, and it was focused on the behavioral competencies associated with highly effective teams. Two primary hypotheses guided this work: (1) operating room team member participants would perceive the simulated scenarios as authentic and worthwhile, and (2) participants would perceive the interdisciplinary simulation training as an effective method for enhancing communication and teamwork during an operating room crisis. Three interdisciplinary operating room teams participated in the training model. Each team completed the training in a virtual minimally invasive operating room using a standardized operative scenario. Each team participated in a debriefing activity immediately after completion of the scenario. An important component included observation of the video-taped performance. Guided by the trainer/facilitator, the team used the videos to identify and discuss specific performances regarding teamwork, communication, and situational awareness. Upon completion of the training, participants completed a voluntary and anonymous training effectiveness questionnaire. Across the 3 teams, all participants felt that the training was worthwhile and would participate again. Eighty percent felt that the session would change their practices in the operating room. All participants felt that the session was effective in promoting team communication skills and crisis-related teamwork. Ninety percent felt that the training model was effective in recognizing operating room errors.

CRM techniques have also been initiated in ED settings, in which similar stressful, dynamic, and

high-stakes environments exist. Many contextual factors characterize the ED and affect how healthcare providers make decisions. These include unclear diagnoses, incomplete or conflicting information, dynamic situations, multiple competing goals, intense time pressure, and serious consequences of error. Three additional CRM principles take these factors into account: triage/prioritization, efficient management of multiple patients, and effective coping with disruptions/distractions.²⁴ Courses have focused on teamwork curricula, training, and refresher training and simulation. Research in ED settings has shown that teamwork skills help decrease clinical errors in the ED.¹⁹

A specific program, the MedTeams team coordination course, was designed by Dynamic Research Corp. (a corporate trainer), which adapted CRM principles from aviation to emergency medicine. It encourages team members to actively coordinate and support one another in the course of clinical task execution by using the structure of work teams. This program operates on the principle that teams and teamwork behavior do not replace clinical skills but instead teamwork actions ensure that clinical activities are properly integrated and executed to deliver effective emergency care.²⁵ In the MedTeams system, a work team, referred to as an ED core team, is a set of 3 to 10 clinically skilled caregivers who work together during a shift and have been trained to use specific teamwork behaviors to tightly coordinate and manage their clinical actions. The team contains at least 1 physician and 1 nurse, and the most experienced physician on the team is designated as the team leader. The team leader serves as leader to only 1 team during a work shift. Team members always know who is on their team and who the team leader is. The team training is organized around 5 team objectives: establishing and maintaining a team structure; planning and problem solving; communicating with the team; managing work load and resources; and improving team skills by review, debriefing, coaching, and teaching.²⁵ The team coordination course has now been adapted for use in labor and delivery units in an attempt to decrease adverse outcomes and improve the process of care in this setting as well.

Neonatal resuscitation has many similarities to medical crises in the ED or operating room. Neonatal resuscitations are often unpredictable, and teams from multiple departments must coordinate and at times compete with little or no preparation. The Neonatal Resuscitation Program (NRP) is the standard curriculum used to teach caregivers how to treat newborns in the delivery room. Thomas *et al.*²⁶ from the University of Texas observed that approximately

30% of NRP steps are not performed or are performed incorrectly. They reported that neonatal resuscitation should be a team activity and that breakdowns in teamwork may contribute to some of the quality problems that are observed. This group conducted a study with 2 specific aims: to incorporate teamwork skills and information about human error into the 1-day NRP training program for interns and to randomize interns to this new version of NRP or the standard NRP and measure the effects on teamwork during the simulated resuscitation following the program. Interns who were randomized to an NRP course with a teamwork and human error curriculum exhibited more team behaviors during the simulated resuscitation than interns in the standard NRP course. This is one of the first studies to document that team training for neonatal resuscitation, and perhaps any healthcare process, can result in more frequent utilization of team behaviors. Other strengths include the random allocation of subjects to the intervention and control groups; the use of trained, blinded observers to rate the frequency of team behaviors; and teaching and measurement of team behaviors that were developed on the basis of research in commercial aviation and neonatal resuscitation. One of the important highlights of this study is that the training curriculum was firmly grounded in the context and processes of neonatal resuscitation. The authors suggest that this grounding for a team training program in specific healthcare processes will be more effective than using generic team training processes and skills.

Teamwork among nursing staff is another area to which attention has been directed. Kalisch *et al.*²⁷ studied a series of interventions to enhance teamwork and staff engagement in a medical unit in an acute care hospital. The interventions involved focus group discussions of perceptions of the level of teamwork in the unit and issues that inhibit and enhance teamwork. These focus group data were compiled into a report that was presented in several feedback sessions. The next step was to conduct sessions to discuss values, visions, and goals, and then a gap analysis was conducted of where they were and where they wanted to be. Each staff member then attended a team training program. Guiding teams were then created to address the specific project goals, and rapid testing and implementation of ideas ensued. The results showed that the intervention resulted in a significantly lower patient fall rate, staff ratings of improved teamwork on the unit, and lower staff turnover and vacancy rates. Improvements in patient satisfaction ratings approached, but did not reach, statistical significance.

TEAMWORK TRAINING IN MEDICAL EDUCATION

There is some existing literature, although limited, concerning how to teach about teams in medical education. Hall and Weaver²⁸ performed a review of the literature on interdisciplinary education and teamwork in healthcare to discover the major issues and best practices. Two main categories emerged from their search: one pertaining to the system of education and training of healthcare professionals and the other pertaining to the content of interdisciplinary education. Several studies have suggested that learning to work in an interdisciplinary milieu should occur early in the education of the healthcare professional (eg, within the first 2 years) because experience of shared learning at an early stage may better facilitate interdisciplinary collaboration. Others have cautioned, however, that one must first be secure in the competency of one's discipline and that a thorough knowledge of one's own discipline is essential to understanding contributions to the team effort.²⁸ Problem-based learning (PBL) is an example of a team-based method of learning. PBL, done in a small-group format, offers a means of integrating theory with clinical components and demands knowledge acquisition and attitudinal and psychomotor skills from its learners. PBL is a structured educational activity, usually employing case presentations as the stimulus to learning. This approach helps students learn to listen to one another and to collaborate as they work to resolve the problem.²⁸ The service/learner model is another teaching method that uses a clinical setting to challenge the learners to work together effectively to address real clinical problems, patient education, or health dilemmas of underserved populations. Both of these methods place the patient at the center of the team's focus.²⁸ Other important concepts addressed in these articles include the need for motivating and educating faculty to learn new tools of team training and the need for the administrative support of the educational and clinical institutions involved in changes.²⁸

Team-based learning (TBL) has become another commonly used curriculum tool. Team learning is a well-defined instructional strategy developed by Larry Michaelson.²⁹ The TBL method allows a single instructor to teach by conducting multiple small groups simultaneously in the same classroom. Learners must actively participate in and out of the class through preparation and group discussion. Class time is shifted away from learning facts and toward the application and integration of information. The

instructor retains control of the content and acts as both facilitator and content expert. The team learning method affords the opportunity for assessment of both individual and team performance. Learners first study independently outside of class to master identified objectives. The individual learners then complete a multiple-choice examination to ensure their readiness to apply their knowledge. Groups of 6 to 7 learners then retake the examination and turn in their consensus answers. Groups then complete class assignments that promote collaboration, use of their acquired knowledge, and identification of learning deficiencies. At designated times, all groups simultaneously share their answers with the entire class for comparison and feedback. This stimulates a large group discussion, with groups defending their answers and the teacher helping to consolidate learning. TBL emphasizes the importance of individual and group accountability, the need and opportunity for group interaction, and the motivation to engage in give-and-take discussion.²⁹ A number of studies that have appeared in the medical literature provide empirical evidence of potential benefits from TBL. Such benefits include increased student engagement, higher quality communication processes, and increased National Board of Medical Examiners shelf examination scores.³⁰

There is little literature about other medical school curricula that teach team training. Charkraborti *et al.*³¹ conducted a systematic review of teamwork training interventions in medical student and resident education. They found 13 studies by searching multiple databases. Most curricula that exist place medical trainees in multidisciplinary learning environments that include nurses, social workers, physical/occupational therapists, administrators, and pharmacists. All curricula employed some active learning methods; these included critical incident simulations, role playing, case-based scenarios, and actual patient encounters. No curriculum described using standardized patients in teamwork training. Three of 13 curricula employed nonmedical teambuilding exercises, such as a rope course or a survival game. The majority of the curricula incorporated feedback as an integral part of active learning. Seven used formal debriefing sessions to provide feedback for learners in the curriculum. Five used facilitated reflection as a technique to help learners gain a richer understanding of teamwork. Three studies used a controlled pre/post evaluation design. Ten used an uncontrolled pre/post design. There were no randomized controlled trials. Five studies were multi-institutional.³¹ Four of the curricula used self-reported measures of knowledge, and none used an objective test of knowledge. None of the knowledge assessments that were described

had been previously validated. Most commonly, knowledge outcomes demonstrated improvements in group awareness. Eight curricula showed improvements in self-assessed team skills. Four curricula included observer assessments of team skills. Nine of the curricula assessed attitudinal changes: all were self-assessed. The reviewers concluded that all of the teamwork curricula analyzed used reasonable educational strategies and appeared to be modestly effective in the short term.³¹

The Agency for Healthcare Research and Quality and the Department of Defense have collaborated to explore the field of medical teamwork. The 2 agencies have developed a new resource training tool for training healthcare providers in better teamwork practices: Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS).³² TeamSTEPPS offers a flexible, evidence-based toolkit to improve patient safety through enhanced communication and other teamwork skills. TeamSTEPPS initiatives occur in 3 continuous phases: assessment; planning, training, and implementation; and reinforcement and sustainment. TeamSTEPPS promotes competency in 4 course areas: team leadership, situation monitoring, mutual support, and communication.³² TeamSTEPPS provides an environment of open communication in which everyone, including the patient, is made to feel comfortable speaking up, regardless of the situation or the complexity of the procedure.³²

The AAMC has implemented the TeamWorks! program, which guides faculty and staff in health professions through approaches to improve the effectiveness of team dynamics and task management. TeamWorks! scholars and faculty coaches provide consultation and support for adapting team behaviors to institutional work.³³

Interprofessional education (IPE) occurs when 2 or more professions learn with, from, and about each other in order to improve collaboration and the quality of practice. IPE helps students to look at a task from the perspective of other professions as well as their own. It enables students to acquire knowledge, skills, and attitudes that they could not acquire in a uniprofessional education.³⁴

Pawlina *et al.*³⁵ at the Mayo Clinic College of Medicine aimed to provide an IPE opportunity to first-year physical therapy and medical students by combining a portion of their respective anatomy laboratory and lecture sessions. Faculty from both disciplines were used, as were formal and informal student interactions between the disciplines. Students from both programs recognized the importance of teamwork across disciplines and appreciated the opportunities to learn from one another and to share

the responsibilities for patient care. This study is a good indication that early introduction to teamwork before the clinical setting is beneficial and well accepted by the students.

Hallin *et al.*³⁴ evaluated whether students perceived that they had achieved interprofessional competence after participating in clinical teamwork training. Six hundred sixteen students from 4 undergraduate educational programs—medicine, nursing, physical therapy, and occupational therapy—participated in an interprofessional course. The interprofessional learning took place in a real clinical work place. The interprofessional student teams collaborated in taking care of and treating patients. The students performed all medical, nursing, physical therapy, and occupational therapy work and care of the patients. The students filled out prequestionnaires and postquestionnaires. All student groups perceived that they had increased their knowledge of the other 3 professions' work and assessed that the course had contributed to the understanding of the importance of communication and teamwork in patient care. All student groups perceived that the clarity of their own professional role had increased significantly. The conclusion of this study was that active patient learning by working together in a real ward context seems to be an effective means to increase collaborative and professional competence.

Debate has ensued about the proper timing of the introduction of IPE. Conventional wisdom has suggested that IPE is most effective in fostering teamwork when introduced to professionals who have a clear sense of their own professional identity and have experiences to share. The opposite viewpoint, however, is that IPE should be introduced at the very start of professional education to prevent the formation of negative interprofessional attitudes that will later be resistant to change.³⁶ Coster *et al.*³⁶ studied students in 8 healthcare groups from 3 higher education institutions in the United Kingdom at 4 time points. They found that the strength of professional identity in all professional groups was high on entry to university but declined significantly over time for some disciplines. Similarly, students' readiness for interprofessional learning was high at entry but declined significantly over time for all groups, except nursing students. They concluded that their findings provide support for introducing IPE at the start of the healthcare students' professional education in order to capitalize on their readiness for interprofessional learning and their well-formed professional identities. However, their study also suggested that students who enter with negative attitudes toward interprofessional learning may gain the least from IPE courses and that an unrewarding

experience in such courses may further reinforce their negative attitudes.

The evaluation of teamwork training is a challenging task for a number of reasons. First, to evaluate the effectiveness of teamwork skills training, it is necessary to measure both teamwork skills exhibited during patient care and team performance in terms of patient outcome. Then, to show that teamwork training improves patient safety, there must be an agreement that the performance measures used are valid proxies for patient safety. In addition, if assessment is done in a simulated environment, it is necessary to demonstrate that the training transfers to clinical situations. Finally, the evaluation must take into account individual and organizational factors that affect training effectiveness.³⁷

There are few validated tools to assess whether or not individuals possess team coordination skills. However, it is imperative to establish a relationship between ratings of team skills and clinical team performance for 2 reasons. First, it is important to empirically show that teamwork skills affect clinical performance and ultimately patient outcomes. Second, it is important to establish the validity of using behaviorally anchored team skill rating scales for assessing individual team skills and for assessing the effectiveness of training programs and interventions.³⁸ Wright *et al.*³⁸ chose to evaluate a rating scale that assessed general team coordination skills that could be applied in a wide variety of healthcare environments. The teamwork dimensions were assertiveness, decision making, situation assessment, leadership, and communication. They recruited 35 medical students to participate in 2 different types of educational tasks on 2 different occasions. Team assignments were random. Each team performed 2 patient assessment tasks in a classroom-based environment. The students used the problem-based teaching method. In the second part of the study, the teams were reassigned and performed 2 high-fidelity simulated patient assessment and care tasks. Two behavioral scientists used a behaviorally anchored team skill rating scale and rated each participant in all 4 cases. In general, individuals were rated higher in teamwork skills in the simulated care tasks than in the classroom tasks. For the classroom task, there was no evidence of a relationship between behaviorally anchored skill ratings and objective team performance measures. In the high-fidelity simulated tasks, there was evidence of a positive relationship between behaviorally anchored skill ratings and objective performance measures. These results may suggest that team skill development requires a dynamic interactive context. The results also provide some support for the validity and reliability of observer-based rating of teamwork skills

when a structured tool is provided for basing ratings on specific observable teamwork behaviors.

Faculty attitudes are believed to be a barrier to successful implementation of IPE initiatives within academic health settings. Curran *et al.*³⁹ studied these attitudes. They distributed a survey to all faculty members in the medicine, nursing, pharmacy, and social work programs in their institution. Respondents were asked to rate their attitudes toward interprofessional healthcare teams, IPE, and interprofessional learning in an academic setting using scales adopted from peer-reviewed literature. Medicine faculty members reported significantly lower scores than nursing faculty in all aspects of the survey. Female faculty and faculty who reported previous experience in IPE reported significantly higher scores. Age, years of practice, and experience as a health professional educator were not related to the overall attitudinal responses. They concluded that faculty development efforts aimed at changing attitudes and increasing understanding of interprofessional collaboration are critical.

CONCLUSION

The practice of medicine is complex. No one individual can expect to care for a patient on his own, and he must be able to effectively interact with other disciplines and specialties to optimize care. Communication across multiple units, physicians, nurses, and others becomes vital to ensuring that accurate and complete information is available, properly exchanged, and regularly updated. Medical education has often emphasized autonomy, and until recently, education on teamwork has not been explicitly included in the medical curriculum. In addition, medicine has traditionally been very hierarchical, emphasizing a communication structure that follows a chain of command philosophy. The IOM explained in its 1999 report that healthcare providers tend to be trained as individuals yet also function exclusively as teams; this creates a gap between training and reality. Teamwork has been studied at length by organizational and human factors psychologists, and as Salas *et al.*¹² stated, "improved teamwork—the seamless integration of multiple knowledge, skill and affective competencies—is a mechanism to improve patient safety and avoid errors." A well-functioning team should be able to do things more effectively and safely than an individual or a group of individuals. A poorly functioning team, however, may be antagonistic and detrimental to productivity.

Teaching teamwork is increasingly relevant in medical education as many factors threaten the ability of trainees to coordinate their individual efforts to care for patients. Medical training places coworkers together for variable and short periods of time. Teams form and then disband. In addition, work-hour restrictions necessitate increasing emphasis on the coordination of roles, facilitation of transfers, and supervision.

There appears to be a consensus in the literature on the need for education about teams in the healthcare professional's education. It is imperative to continue to develop educational curricula and training models for teamwork and collaborative practice across the continuum of professional education and to implement educational initiatives about the new model of interprofessional healthcare delivery and teamwork. Certain nonclinical skills must be learned by team members: group skills, communication skills, and conflict resolution skills. It is also vital to develop leadership training opportunities and to foster a culture that supports teamwork and builds the capacity for team members to practice collaboratively. Faculty development must be provided to address motivations to participate. This includes opportunities for faculty to learn how to facilitate interdisciplinary education sessions and to learn skills from management and behavioral sciences. By drawing on the science of teams and the work on team training that has been done in other fields, we can focus our endeavors on applying these methods to a medical curriculum that will encompass undergraduate and graduate training and accomplish the charge delivered by Kirch and the IOM. As Salas⁴⁰ recently stated in his presentation at the 2008 AAMC meeting, "team training and teamwork must be part of the DNA of healthcare."⁴⁰

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Kohn L, Corrigan J, Donaldson MS. *To Err Is Human: Building a Safer Health Care System*. Washington, DC: National Academy Press; 2000.
2. Oandasan I, Baker G, Barker K, et al. Teamwork in healthcare: promoting effective teamwork in healthcare in Canada. Available at: <http://www.chsrf.ca/research-themes/pdf/teamwork-synthesis-report.e.pdf>. Accessed May 2009.
3. Agency for Healthcare Research and Quality. Medical teamwork and patient safety. Chapter 5. Conclusions and recommendations. Available at: <http://www.ahrq.gov/qual/medteam/medteam5.htm>. Accessed May 2009.

4. Kirch DG. AAMC president's address 2007: culture and the courage to change. Available at: <http://www.aamc.org/meetings/annual/2007/highlights/president.htm>. Accessed May 2009.
5. Katzenbach JR, Smith DK. *The Wisdom of Teams: Creating the High-Performance Organization*. New York, NY: Harper Business; 1993.
6. Salas E, Cannon-Bowers JA. Teamwork and team training. In: Smelser NJ, Baltes PB, eds. *International Encyclopedia of the Social and Behavioral Sciences*. Oxford, England: Elsevier Science; 2001; 15:487–15:492.
7. Farley M. Teamwork in perioperative nursing. *AORN J* 1991; 53: 730–738.
8. Wells C. Teaching teamwork in information systems. In: Cohen E, ed. *Challenges of Information Technology Education in the 21st Century*. Hershey, PA: IGI Publishing; 2002; 1–24.
9. Katzenbach JR, Smith DK. The discipline of teams. *Harvard Bus Rev* 1993; the best of 1993: 111.
10. Kirkman B, Rosen B. Powering up teams. *Org Dyn* 2000; 28: 48–65.
11. Busseri M, Palmer J. Improving teamwork: the effect of self-assessment on construction design teams. *Des Stud* 2000; 21: 223–238.
12. Salas E, Sims DE, Klein C, Burke CS. Can teamwork enhance patient safety? *Forum Risk Manage Found Harvard Med Inst* 2003; 23: 5–9.
13. Edmonson A, Bohmer R, Pisano G. Speeding up team learning. *Harvard Bus Rev* 2001; 125. October 2001.
14. Accreditation Council on Graduate Medical Education. Outcome project: assessing residency education through outcomes assessment. Available at: <http://www.acgme.org/outcome>. Accessed May 2009.
15. *Contemporary Issues in Medicine: Communication in Medicine. Medical School Objectives Project*. Washington, DC: Association of American Medical Colleges; 1999.
16. Musson DM, Helmreich RL. Team training and resource management in health care: current issues and future directions. *Harvard Health Policy Rev* 2004; 5: 25–35.
17. Sundar E, Sundar S, Pawloski J. Crew resource management and team training. *Anesthesiol Clin* 2007; 25: 283–300.
18. Hunt E, Shilkofski N, Stavroudis T, Nelson K. Simulation: translation to improved team performance. *Anesthesiol Clin* 2007; 25: 301–319.
19. McConaughy E. Crew resource management in healthcare. The evolution of teamwork training and MedTeams. *J Perinat Neonatal Nurs* 2008; 22: 96–104.
20. Agency for Healthcare Research and Quality. Medical teamwork and patient safety. Chapter 3. Team training in high-risk contexts. Available at: <http://www.ahrq.gov/qual/medteam/medteam3.htm>. Accessed May 2009.
21. Crichton M. Attitudes to teamwork, leadership, and stress in oil industry drilling teams. *Saf Sci* 2005; 43: 679–696.
22. Gaba D, Howard S, Fish K, et al. Simulation-based training in Anesthesia Crisis Resource Management (ACRM): a decade of experience. *Simul Gaming* 2001; 32: 175–193.
23. Paige J, Kozmenko V, Morgan B, et al. From the flight deck to the operating room: an initial pilot study of the feasibility and potential impact of true interdisciplinary team training using high fidelity simulation. *J Surg Educ* 2007; 64: 369–377.
24. Eppich W, Brannen M, Hunt E. Team training: implications for emergency and critical care pediatrics. *Curr Opin Pediatr* 2008; 20: 255–260.
25. Risser D, Rice M, Salisbury M, et al. The potential for improved teamwork to reduce medical errors in the emergency department. *Ann Emerg Med* 1999; 34: 373–383.
26. Thomas E, Taggart B, Crandell S, et al. Teaching teamwork during the neonatal resuscitation program: a randomized trial. *J Perinatol* 2007; 27: 409–414.
27. Kalisch B, Curley M, Stefanov S. An intervention to enhance nursing staff teamwork and engagement. *J Nurs Adm* 2007; 37: 77–84.
28. Hall P, Weaver L. Interdisciplinary education and teamwork: a long and winding road. *Med Educ* 2001; 35: 867–875.
29. The Team-Based Learning Collaborative. Available at: <http://tblcollaborative.org>. Accessed May 2009.
30. Thompson B, Schneider V, Haidet P, et al. Team-based learning at ten medical schools: two years later. *Med Educ* 2007; 41: 250–257.
31. Chakraborti C, Goonyasai R, Wright S, Kern D. A systematic review of teamwork training interventions in medical student and resident education. *J Gen Intern Med* 2008; 23: 846–853.
32. Clancy C, Tornberg D. TeamSTEPPS: assuring optimal teamwork in clinical settings. *Am J Med Qual* 2007; 22: 214–217.
33. Grigsby RK, Magrane D. Teams as tools for changing the culture of academic medicine. Available at: <http://www.aamc.org/members/facultydev/faculty-vitae/summer08/feature.htm>. Accessed May 2009.
34. Hallin K, Kiessling A, Waldner A, Henriksson P. Active interpersonal education in a patient based setting increases perceived collaborative and professional competence. *Med Teach* 2009; 31: 151–157.
35. Hamilton SS, Yuan BJ, Lachman N, et al. Interprofessional education in gross anatomy: experience with first-year medical and physical therapy students at Mayo Clinic. *Anat Sci Educ* 2008; 1: 258–263.
36. Coster S, Norman I, Murrells T, et al. Interprofessional attitudes amongst undergraduate students in the health professions: a longitudinal questionnaire survey. *Int J Nurs Stud* 2008; 45: 1667–1681.
37. Entin E, Lai F, Barach P. Training teams for the perioperative environment: a research agenda. *Surg Innov* 2006; 13: 170–178.
38. Wright M, Phillips-Bute B, Petrusa E. Assessing teamwork in medical education and practice: relating behavioral teamwork ratings and clinical performance. *Med Teach* 2009; 31: 30–38.
39. Curran V, Sharpe D, Forristall J. Attitudes of health sciences faculty members towards interprofessional teamwork and education. *Med Educ* 2007; 41: 892–896.
40. Salas E. Does team training work? Insights and warnings for health care. Paper presented at: National AAMC Meeting; 2008; San Antonio, TX.

The Utility of Simulation in Medical Education: What Is the Evidence?

Yasuharu Okuda, MD,^{1,2} Ethan O. Bryson, MD,¹ Samuel DeMaria Jr, MD,¹ Lisa Jacobson, MD,¹ Joshua Quinones, MD,³ Bing Shen, MD,¹ and Adam I. Levine, MD¹

¹Mount Sinai School of Medicine, New York, NY

²Institute for Medical Simulation and Advanced Learning, Health and Hospitals Corporation, New York, NY

³Keck School of Medicine, University of Southern California, Los Angeles, CA

ABSTRACT

Medical schools and residencies are currently facing a shift in their teaching paradigm. The increasing amount of medical information and research makes it difficult for medical education to stay current in its curriculum. As patients become increasingly concerned that students and residents are “practicing” on them, clinical medicine is becoming focused more on patient safety and quality than on bedside teaching and education. Educators have faced these challenges by restructuring curricula, developing small-group sessions, and increasing self-directed learning and independent research. Nevertheless, a disconnect still exists between the classroom and the clinical environment. Many students feel that they are inadequately trained in history taking, physical examination, diagnosis, and management. Medical simulation has been proposed as a technique to bridge this educational gap. This article reviews the evidence for the utility of simulation in medical education. We conducted a MEDLINE search of original articles and review articles related to simulation in education with key words such as *simulation*, *mannequin simulator*, *partial task simulator*, *graduate medical education*, *undergraduate medical education*, and *continuing medical education*. Articles, related to undergraduate medical education, graduate medical education, and continuing medical education were used in the review. One hundred thirteen articles were included in this review. Simulation-based

training was demonstrated to lead to clinical improvement in 2 areas of simulation research. Residents trained on laparoscopic surgery simulators showed improvement in procedural performance in the operating room. The other study showed that residents trained on simulators were more likely to adhere to the advanced cardiac life support protocol than those who received standard training for cardiac arrest patients. In other areas of medical training, simulation has been demonstrated to lead to improvements in medical knowledge, comfort in procedures, and improvements in performance during retesting in simulated scenarios. Simulation has also been shown to be a reliable tool for assessing learners and for teaching topics such as teamwork and communication. Only a few studies have shown direct improvements in clinical outcomes from the use of simulation for training. Multiple studies have demonstrated the effectiveness of simulation in the teaching of basic science and clinical knowledge, procedural skills, teamwork, and communication as well as assessment at the undergraduate and graduate medical education levels. As simulation becomes increasingly prevalent in medical school and resident education, more studies are needed to see if simulation training improves patient outcomes. *Mt Sinai J Med* 76:330–343, 2009. © 2009 Mount Sinai School of Medicine

Key Words: graduate medical education, simulation, undergraduate medical education.

Address Correspondence to:

Yasuharu Okuda, MD

Departments of

Emergency Medicine and

Medical Education

Mount Sinai School of Medicine

New York NY

Email: yasuharu.okuda@mssm.edu

As educators in the field of medicine, we have a tremendous responsibility to our students, patients, and society as a whole. Participating in medical students' professional development is truly a gratifying yet daunting task. We must help our students to develop the ability to recognize their own limitations and knowledge gaps and provide them with the tools to fill these voids. Simulation, in all its incarnations, is a tremendous tool for healthcare educators, in that it allows students to achieve these goals without our patients being put at risk. Simulation, as defined by

Dr. David Gaba,¹ is an instructional process that substitutes real patient encounters with artificial models, live actors, or virtual reality patients. The goal of simulation is to replicate patient care scenarios in a realistic environment for the purposes of feedback and assessment. Properly conducted, simulation creates an ideal educational environment because learning activities can be made to be predictable, consistent, standardized, safe, and reproducible. This environment encourages learning through experimentation and trial and error with the ability to rewind, rehearse, and practice without negative patient outcomes. It is not our intent to suggest that simulation, in its present form, serves as an acceptable substitute for the actual clinical experience, and we agree that the ideal setting for clinical education remains the actual clinical environment, but unfortunately, the ideal is not always practical.

One might question the necessity of simulation in medical education when the apprenticeship model of training under experienced clinicians has served us well for so many years. However, the world of medicine is changing because of a variety of issues affecting medical education. Patients are now increasingly treated on an outpatient basis, and less time is afforded to patient interaction. Continuity of care has also decreased as time limits are placed on resident work hours. Simulation technology has begun to gain widespread acceptance in medical education because of the safety of the environment, the ability to demonstrate multiple patient problems, the reproducibility of content, and the ease of simulating critical events.² Coupled with an aggressive medicolegal system in the United States, the idea of learning difficult and error-prone tasks by making errors on live patients is progressively less acceptable.³ It has been said that we remember from our failures, not our successes. Those events that end with an untoward outcome create "seasoned" veteran clinicians.

To date, there is a small but growing body of evidence that simulator training improves healthcare education, practice, and patient safety, but Gaba⁴ argues that "no industry in which human lives depend on skilled performance has waited for unequivocal proof of the benefits of simulation before embracing it." Patient safety and medical errors have come to the forefront of healthcare since the Institute of Medicine released *To Err Is Human: Building a Safer Health System* in 2000.⁵ The effective integration of simulation into medical education and assessment can address this modern healthcare challenge.

Medical simulation is often divided into 4 areas by the educational tool: a standardized patient, a screen-based computer, a partial-task simulator, and

a high-fidelity mannequin simulator.⁶ This article reviews the utility of simulation in medical education and focuses mainly on the latter 3 areas related to simulation technologies, as a vast body of literature already exists for standardized patients.⁷ We briefly discuss the history of mannequin simulation and types of simulators and then focus on the educational theory behind simulation as well as simulation as it applies to undergraduate medical education, graduate medical education, and continuing medical education (CME).

HISTORY OF THE MANNEQUIN SIMULATOR

The first mannequin used to teach airway and resuscitative skills was developed by 2 anesthesiologists, Dr. Peter Safar, an American, and Dr. Bjorn Lind, a Norwegian, during the 1950s.⁸ Lind worked with a toy manufacturer to develop what has been known for over half a century as Resusci-Annie. Some 10 years later, Dr. Stephen Abrahamson, also from the United States, presented the advantages of training anesthesiologists with his full-scale, computer-controlled human patient simulator.⁹ Toward the end of the 1980s, 2 teams of anesthesiologists, one from the University of Florida (Dr. Michael Good and Dr. John Gravenstein) and the other from Stanford (led by Gaba), developed a realistic mannequin simulation. They combined engineering skills and the idea of using simulation in education for team training and with the aim of improving patient safety. The result was an interactive, realistic patient simulator that could replicate the human response to various physiological and pharmacological interactions.¹⁰ At the same time, Dr. Hans-Gerhard Schaefer and his colleagues in Basel, Switzerland developed full-scale simulators for crisis resource training of operating room teams.¹¹ Today, these high-fidelity mannequin simulators are capable of recreating physical examination findings, including normal and abnormal heart and lung sounds, pupil diameter, sweating, and cyanosis, as well as physiological changes, such as changes in blood pressure, heart rate, and breathing.

TYPES OF SIMULATIONS

Simulation, as defined earlier, is the replacement of real patient encounters with either standardized patients or technologies that replicate the clinical scenario. These technologies range from simple demonstrations through screen-based video game-type

simulations, such as fiber-optic bronchoscopy, to partial-task trainer devices designed to allow learners to practice procedures, including chest tube insertion and intubation, to full environment simulation (FES) using high-fidelity mannequin simulators.

Flight simulators are perhaps the most highly developed form of FES: full cockpit environments undergo all of the movements associated with real flight. High-fidelity mannequin simulators are typically used to recreate real patient encounters in a simulated clinical environment. Even the highly advanced mannequin simulators in use today can be used along a continuum of fidelity, reflecting the time and resources available to the instructor. Students participating in an anesthesiology-based curriculum may be run through a simple anesthetic induction and evaluated on their performance with a checklist of required tasks in a particular order. On the other end of the spectrum is the simulation theater, in which details, both subtle and overt, are embedded, creating emotion, confusion, and distraction and adding that element of reality which helps to enhance the experience. Instead of simply completing the tasks on an anesthesia checklist (eg, start an intravenous line, apply monitors, pre-oxygenate, induce anesthesia, and intubate), participants may face a difficult or inappropriate patient, may have to deal with an obnoxious or verbally abusive family member, or may have to address an inappropriate comment from a consultant. Standardized patients, played by actors, can be used in conjunction with FES to lend an even greater level of reality to simulation. This theater, which adds so much to the reality of the simulation experience, can be laced with emotional content to enhance the learning experience.

The ingredients of a full-environment simulation include the patient (high-fidelity mannequin simulator), other healthcare professionals, and ancillary equipment and supplies designed to replicate the clinical environment. Current mannequin-based simulator designs are computer- and model-driven, full-sized infant, child, or adult patient replicas that are capable of delivering "true-to-life" scenarios that simulate reality. The incorporation of the mannequins into a simulated clinical environment complete with monitors and medical equipment commonly found in real clinical scenarios allows participants the ability to suspend disbelief, thus creating a highly effective learning environment.¹² The lessons learned can be perceived as more realistic in comparison with simpler screen-based simulators.

FES provides the unique opportunity to not only practice procedures but also allow educators the ability to stage realistic settings in which the principal focus can be human behavior and

interaction. In this environment, participants can be allowed to make mistakes and experience bad outcomes without patient harm.¹³ These adverse outcomes can facilitate the generation of negative emotions among the participants, as no healthcare professional wants to be responsible for contributing to patient harm through a bad clinical outcome, especially when witnessed by colleagues. One of the major manufacturers alleges that its device "exhibits clinical signals so lifelike that students have been known to cry when it 'dies.'"¹² We theorize that the effectiveness of the simulator as an educational tool not only depends on the ability of the simulator to realistically emulate human physiology and physiological responses but also depends on the specially designed facilities and the expertise of the educators to accomplish FES that triggers these emotions.

SIMULATION AS AN EDUCATIONAL TOOL

The medical learner at any stage—undergraduate, graduate, or postgraduate—is truly an adult learner. Defined in various ways by education theorists, the adult learner, by all accounts, is seen to learn by different methods and for different reasons in comparison with earlier stages in his education. In 2008, building on the work of other education scholars (including Knowles, the father of andragogy, or adult learning theory), Bryan *et al.*¹⁴ described 5 adult learning principles that apply to the medical learner (Table 1).

These are an adaptation of the assumptions of andragogy, which emphasizes the role of experience and self-direction as well as the need to know the benefits of knowledge and its potential applications before one embarks on the instructional journey.¹⁵ The challenge then is to create effective educational pathways within this paradigm. The complexity

Table 1. Five Adult Learning Principles That Apply to the Medical Learner.

1. Adult learners need to know why they are learning.
2. Adult learners are motivated by the need to solve problems.
3. The previous experiences of adult learners must be respected and built upon.
4. The educational approach should match the diversity and background of adult learners.
5. Adults need to be involved actively in the process.

NOTE: The material for this table was taken from Bryan *et al.*¹⁴

grows, however, when the technical aspects of clinical instruction are introduced. Clinical skills, on the simplest level, are psychomotor skills learned via reinforcement and requiring straightforward instruction. In some ways, clinical skills would seem easier to master because students tend to remember 90% of what they do yet only 10% of what they read.¹⁶ However, clinical skills consist of more than motor memory. They require the integration of problem-solving skills, communication skills, and technical skills in the setting of a complex medical context, and they require practice.

The ultimate goal in medical education is expertise or mastery of one's trade. Deliberate practice is an educational technique used to produce expert performance contingent upon 4 conditions: intense repetition of a skill, rigorous assessment of that performance, specific informative feedback, and improved performance in a controlled setting.¹⁷ The true goal of deliberate practice, similar to the true goal of medical education, is to produce an expert or master of any trade. As Eppich *et al.*¹⁸ explained, unless the goal is improved performance, merely participating in the activity will not lead to mastery. It would appear then that one way to improve medical education is to seek a mechanism for deliberate practice. A recent study by Wayne *et al.*^{19,20} assessed the value of using simulation technology and deliberate practice to enhance acute resuscitation skills. This study of 41 second-year internal medicine residents using deliberate practice to teach advanced cardiac life support (ACLS) showed statistically significant improvements in education outcomes, including compliance with standard ACLS protocols as well as retention of skills and knowledge after 14 months.

In evaluating simulation as an educational tool, Issenberg *et al.*²¹ and McGaghie *et al.*²² described the features enhancing its utility (Table 2). Simulation provides the tools and the paradigm to enhance medical education, but it depends on intense preparation and support from faculty as well as buy-in from the participants. The ability to provide immediate directed feedback is the primary advantage of simulation. This opportunity is typically lacking in the clinical setting. It also effectively addresses the diversity of both learners and situations with its adaptable, programmable structure. The main limitation of simulation is learner-dependent, as it requires full participation and engagement by the individual.

FES addresses many of the challenges in today's medical education. First, it provides a medium and mechanism for clinical skill instruction. Everything from basic interview skills to lung auscultation

Table 2. *Features and Uses of High-Fidelity Medical Simulations That Lead to Effective Learning*

1.	Mechanism for repetitive practice
2.	Ability to integrate into a curriculum
3.	Ability to alter the degree of difficulty
4.	Ability to capture clinical variation
5.	Ability to practice in a controlled environment
6.	Individualized, active learning
7.	Adaptability to multiple learning strategies
8.	Existence of tangible/measurable outcomes
9.	Use of intra-experience feedback
10.	Validity of simulation as an approximation of clinical practice

NOTE: The material for this table was taken from Issenberg *et al.*²¹ and McGaghie *et al.*²²

to emergent cricothyrotomies can be practiced in a simulated environment. That environment allows for relevant, active learning with feedback, these being important components of clinical skill development.²³ Next, it creates a safe environment for practicing with new technologies and, importantly, ensures that practice occurs without endangering patient or practitioner safety. Traditional practice of "see one, do one, teach one" may no longer be ethical.²⁴ FES is also an ideal environment for the promotion of teamwork and communication. Many clinical problems cannot be solved simply by identification of the diagnosis; they often require resource management and teamwork, 2 practical skills well suited to simulation education.^{25,26} A study by Ottestad *et al.*²⁶ demonstrated that simulation could be used to reliably assess nontechnical skills such as interpersonal communication and task management.¹⁸ In Ottestad *et al.*'s study, house staff were videotaped managing a standardized simulated septic shock patient in teams, and a benchmark checklist of clinical and nonclinical skills was developed. The tapes were reviewed by 2 of the study authors pretrained in skill review, and interrater reliability was found in the assessment of both clinical and nonclinical skills ($r = 0.96$ and $r = 0.88$, respectively). Lastly, training with high-fidelity mannequin simulation improves knowledge consolidation. Wayne *et al.*^{20,27} demonstrated by both retrospective case log review and simulator testing that comparing the acute cardiac arrest management of residents trained by simulation and those not trained on the simulator showed that students learning new material with the use of simulation retained knowledge further into the future than students studying with more traditional methods.

Medical learners may also benefit from experiential learning, a model explained by David Kolb

that emphasizes a process of learning building on concrete experience. His model describes 4 stages: (1) concrete experience, (2) observation and reflection, (3) abstract concept formation, and (4) active experimentation in which generalizations are tested and new hypotheses are developed to be tested in future concrete experiences. The simulation experience affords an excellent opportunity to expand on this model. Learners are thrown into a simulated concrete experience that allows them to progress through the cycle, ideally developing skills and knowledge to be applied in future simulated or actual concrete experiences. One of the most important parts of the experiential learning cycle is debriefing, a process that is often difficult to perform in the typical clinical learning experience. In a simulated environment, debriefing can be successfully accomplished.²⁸ A study by Savoidelli *et al.*²⁸ found simulation without debriefing showed no improvement in nontechnical skills in comparison with simulation with either oral or videotape-assisted oral feedback, which showed significant improvement ($P > 0.005$).²⁸

UNDERGRADUATE MEDICAL EDUCATION

The recent growth of interest in simulation has led to a cascade of applications as a supplement to or replacement for current models of undergraduate medical education. The reason for this is the strong belief in its efficacy as an addition to and sometimes replacement of the current models of education. Traditional modes of education rely on noninteractive classroom lectures and more recently problem-based learning formats to relay basic science concepts and disease processes. Simulation-based training seeks to teach these concepts as well as cognitive and motor skills in an interactive way that more accurately reflects the clinical experience.

Many studies have shown that simulation is a valuable educational tool in undergraduate medical education. Simulation has been used as an evaluation tool to assess knowledge gaps in medical students and residents in the management of acutely ill patients.²⁹ A study by Young *et al.*²⁹ tested medical students and residents in the management of high-risk scenarios and sought to show differences in performance for each level of training with a steady increase in performance as training increased. Surprisingly, when the participants were assessed, all groups performed equally poorly in comparison with the expert-attending comparison group. This performance deficit signified that there might be a

deficiency in the current model of the lecture-based and problem-based learning format of teaching. This also revealed potential dangers in the current medical system in which residents decide many critical actions.

As a teaching tool, simulation-based teaching has proved superior to the traditional problem-based learning format. A study by Steadman *et al.*³⁰ randomized fourth-year medical students to receive a problem-based learning or simulation-based teaching training intervention for the management of acute dyspnea. Each group received the other intervention for the management of abdominal pain as a control. After a weeklong study period that included an initial assessment, intervention, and final assessment on a simulator, scores were tallied on the basis of standardized performance checklists. The results showed that the group receiving the simulation intervention performed significantly better with a greater improvement in scores from baseline than the problem-based learning group. Next is an overview of simulation use within the various levels of medical school education.

Basic Science

Medical student educators have used simulation in basic science education. Online computer-based simulators and, more recently, high-fidelity mannequin simulators have been used to teach physiology and have been well received by both students and educators as an effective educational tool.^{31,32} Via *et al.*³² used a full-scale human simulator to demonstrate changes in cardiac output, heart rate, and systemic vascular resistance in response to volatile anesthetics for second-year medical students learning pharmacology. Student acceptance was high, with 95% of the students feeling that it was a valuable use of their time and 83% of the students preferring this method over didactic lectures. In another study by Seybert *et al.*,³³ students underwent cardiac scenarios including dysrhythmias, hypertensive urgency, and myocardial infarction. On the basis of preintervention and postintervention scores, students demonstrated improvement of knowledge, ability, and self-confidence in pharmacological management. Simulation in basic science education is often conducted in small-group sessions with the reasoning that it is more effective in such settings. However, Fitch³⁴ showed that it was possible to teach large groups basic neuroscience concepts with simulation in large settings. High-fidelity patient encounters were presented with physician actors in front of a large group of first-year and second-year medical students to demonstrate basic neuroscience concepts.

A review of the videotaped large-group sessions showed extensive interactions within the groups, and student feedback revealed that 98% of students felt the correlation to basic science concepts was good or outstanding.

Physical Examination

Simulation has also been used to help teach physical examination skills such as the heart and lung examination. The cardiology patient simulator (CPS), developed by Dr. Michael Gordon in 1968, can be used to teach 30 different cardiac conditions, reproducing abnormal and normal findings such as respiratory sounds, heart sounds, pulses, jugular venous pulsations, and precordial pulsations. This part-task simulator was developed because of the limitations of relying on patient exposure to teach all abnormal cardiovascular examination findings to medical students. A multicenter study, involving 208 medical students, demonstrated significant improvement in clinical skills and knowledge when students were trained on a CPS during a fourth-year cardiology clerkship in comparison with standard bedside teaching.^{6,35} Students on a cardiology clerkship were divided into a control group using standard bedside teaching and a CPS group using both standard and CPS training. Pretests between the 2 groups were not statistically significant. Upon completion of the training, the students in the CPS group showed statistically significant improvement on a written multiple-choice test and a clinical skills examination with both a CPS and a real patient. Issenberg *et al.*^{36,37} showed similar findings in medical house staff.

Clinical Clerkships

The clinical clerkships are a period during which medical students begin to apply preclinical knowledge to the care of real patients. Simulation can help bridge this gap. Medical students have been placed in a variety of simulated scenarios, including airway management in anesthesia cases,³⁸ trauma management,³⁹ and critical care management.⁴⁰ Murray *et al.*³⁹ used trauma simulation to explore measurement properties of scores. Third-year and fourth-year medical students and first-year residents underwent a trauma simulation exercise and were subsequently scored on the basis of item-checklist completion, thought process, action performance, and care prioritization. Results from the study showed that timing and sequencing of actions distinguished high-ability examinees from low-ability examinees and correlated with the training level. This and other studies such as that of Boulet *et al.*⁴⁰ have shown that

simulation is a valid measure of performance.⁴¹ As a tool for performance improvement, Morgan *et al.*⁴² had 299 fourth-year medical students undergo simulation sessions that involved working through 4 clinical scenarios that resulted in an unstable cardiac arrhythmia. Subsequent testing of knowledge and a review of actions showed statistically significant improvement in written test scores, global ratings, and completion of performance checklists.

Skills Training

Simulation-based training has also found utility in procedural and surgical skills training. Medical student training with virtual reality simulation has similarly led to greater knowledge acquisition and improved performances in simulated surgeries.⁴³ In one study by Van Sickle *et al.*,⁴⁴ medical students training with a laparoscopic suturing part-task simulator were able to learn advanced technical skills comparable to those of senior level residents in a short amount of time. For nonsurgical procedures such as cricothyrotomies, central lines, and chest tubes, medical students with simulation training felt more comfortable with and more willing to perform these procedures on their own than those trained without simulation.⁴⁵ In obstetrics, simulation has been used to teach management of shoulder dystocia and simple delivery.⁴⁶ Graber *et al.* demonstrated that patients are more willing to allow medical students to perform procedures such as venipuncture, lumbar puncture, or central lines on them after they have undergone simulation training.⁴⁷

GRADUATE MEDICAL EDUCATION

In 1999, the Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties defined competency in terms of 6 domains: patient care (including clinical reasoning), medical knowledge, practice-based learning (including information management), interpersonal and communication skills, professionalism, and systems-based practice (including health economics and teamwork). During the following decade, the ACGME identified common requirements for all residency programs. These common requirements mandate that programs must educate, evaluate, remediate, and determine that their graduates are competent in terms of these 6 domains. It is expected that the residency programs will document these processes. Because teaching and assessing some of these domains is complex and imperfect, the ACGME furnished a

"tool box" with recommended teaching and evaluation modalities. The ACGME considered the use of simulation, including standardized patients and mechanical simulators, to be extremely effective for teaching and evaluating some of the so-called light competencies such as communication skills, professionalism, and systems-based practice. However, all 6 of the ACGME core competencies can be addressed through simulation.

Anesthesiology

Anesthesiologists such as Dr. David Gaba of Stanford Medical School have led the way in anesthesia simulation since the 1980s. Given the nature of anesthesiologists' practice environment, opportunities for learning are often sporadic. When adverse or crisis events do occur, little teaching takes place, and ubiquitous legal concerns often prevent debriefing and discussion among colleagues. Simulation enables learners to witness such events in a safe environment and creates learning opportunities.

A 1999 survey by Morgan and Cleave-Hogg⁴⁸ found that 71% of medical schools used some form of mannequin or simulator to teach anesthesia to medical students. Eighty percent of these institutions also used simulation for postgraduate training. This widespread use has been prompted by evidence that medical decision making and the occurrence of errors can be effectively examined in simulation labs and that improved performance can result from simulator training.^{49–53} There is also evidence that simulation training improves provider and team competence on mannequins and that procedural simulation improves actual performance in clinical settings with respect to implementing ACLS protocols.⁵⁴

It is not surprising that governing bodies have taken notice. Indeed, the American Board of Anesthesiology now requires simulator-based education to fulfill maintenance of certification in anesthesia requirements,⁵⁵ and Ziv *et al.*⁵⁶ reported the use of simulation by the Israeli Board of Anesthesiology Examination Committee as a crucial element in credentialing and certifying anesthesiologists. It is likely that the role of simulation will only grow in the field of anesthesiology, in which rare events and resource management training are so crucial.

Surgery

The role of simulation in surgical education is rapidly changing the traditional apprenticeship model, which often suffers from poor reliability when applied to performance assessment and teaching.⁵⁷ Surgeons need to learn new procedures and become proficient

with new tools throughout their careers. The acquisition of new knowledge and skills can be especially challenging for surgeons, who often work under tremendous time constraints.^{58–60} Although surgical simulation can be thought to start on the first day of gross anatomy class, high-fidelity, computerized models are replacing and augmenting traditional training.⁶¹

An impressive array of simulators is now available for use in the teaching, learning, and assessment of surgical competencies (from part-task trainers to virtual reality). The continuum of surgical care may be simulated through interconnected stations, each one focusing on a phase in the care of the surgical patient.⁶² A review of surgical simulation showed that computer simulation generally showed better results than no training at all.⁶³ Most surgical trainees feel that simulation is essential to their current surgical training curricula.⁶⁴ Furthermore, data suggest the potential for clinical benefits from simulation-based skills development, especially in laparoscopic⁶⁵ and endoscopic procedures (upper endoscopy and colonoscopy).⁶⁶ Studies by Aggarwal *et al.*,⁶⁷ Stefanidis *et al.*,⁶⁸ and Korndorffer *et al.*⁶⁹ demonstrated that laparoscopic surgery simulation proved effective for surgical residents in training, maintaining skills, and improving performance in the operating room.

Obstetrics

A number of part-task trainers have since been created for training in common tasks (eg, determining the degree of cervical dilation), and high-fidelity birthing simulators now have motor-driven mechanics that can move a mannequin fetus out of the birth canal. The extant literature supports simulation for practicing routine and uncommon procedures and events, for improving technical proficiency, and for fostering self-confidence and teamwork among obstetricians as measured by self-report.^{46,70} Many articles have been published describing simulators for teaching ultrasound-guided amniocentesis,⁷¹ determining fetal station,⁷² managing shoulder dystocia,^{62,73–78} and managing obstetric emergencies and trauma.^{79,80} A recent retrospective study by Draycott *et al.*,⁷⁵ which looked at neonatal outcome before and after shoulder dystocia training on a simulator, showed a reduction in neonatal injury from 9.3% to 2.3%.

Emergency Medicine

The use of simulation in emergency medicine has expanded since the late 1990s. A recent survey of

emergency medicine residencies by Okuda *et al.*⁵ showed that 91% used simulation in training their residents. Of those using simulators, 85% used mannequin-based simulators. Of the emergency medicine residencies using simulation, many of the programs are using simulation to train skills and algorithms such as intubation and resuscitation. To this end, one study demonstrated that simulation-based rapid-response team training correlated with improved team functioning and adherence to American Heart Association guidelines in real in-hospital emergencies.⁸¹

Crew resource management, in particular, is important in the emergency department, and much of the work done with emergency medicine simulation revolves around this concept. Crew resource management was a concept developed by the National Aeronautics and Space Administration to evaluate the role of human factors in high-stress and high-risk environments such as aviation. This model was adapted into medicine as crisis resource management. Currently, Stanford Medical School runs a compulsory crisis management course for its emergency medicine residents, with increasingly complex scenarios presented as residents progress through their training.⁸² Numerous studies have concluded that participants in similar courses viewed their crew resource simulations favorably and believed that their knowledge and skills had been improved.⁸³ In a study by Bond *et al.*,¹³ emergency medicine residents were exposed to scenarios in which they were expected to fail and later examine the decisions that led to their errors. The participants stated that they learned from their errors and ranked the experience second only to actual patient care in terms of educational benefit. Although the current evidence does not conclusively support whether simulation training can change behaviors, skills, or performance in emergency medicine, the increased interest in these possibilities means more research in coming years.

Pediatrics

Although much of the simulation work in general pediatrics has been focused on standardized patient encounters, high-fidelity simulation in pediatrics has focused on neonatal resuscitation. Numerous studies have highlighted deficiencies in providers' conformity to neonatal resuscitation guidelines.^{84,85} Hence, there exists a need for better training in the skills and teamwork involved in handling pediatric resuscitation and emergencies. Halamek and his group at Stanford University⁸⁶ developed a neonatal resuscitation program in the mid-1990s. The

program used high-fidelity pediatric simulation and emphasized teamwork and technical skills. Trainees reported that the simulations enhanced their critical thinking, behavioral, and technical skills. Although data are lacking on the transfer of these skills from the simulator to real patients, it is likely that the confidence gained through simulation augments the knowledge gained through actual patient encounters.

Critical Care

The critical care setting is highly dynamic and necessitates many skills that can be taught or honed through simulation. As in many other fields, teamwork and leadership skills are crucial to critical care and can be rehearsed in a simulation environment.⁸⁷ Procedures such as central line placement can also be practiced. Central line simulation before actual performance on patients was well received by participants in one study.⁸⁸ Still, no long-term outcome evidence exists about whether simulation training improves outcomes.

Competency Assessment and Physician Impairment

Public pressure on physician performance is increasingly compounding the dilemma of assessing competency. Traditionally, physicians were assessed while delivering care to live patients in real clinical situations. This method is growing increasingly unpopular because of public demand for an error-free learning environment. In fact, some malpractice suits have attempted to shift blame to the inadequacies of residency programs.⁸⁹ Bond and Spillane⁹⁰ outlined how simulation could be used to assess the ACGME core competencies in emergency medicine residents. Other work in the field of emergency medicine has elaborated on the evaluation of specific core competencies and ethical dilemmas.^{91,92} This expands assessment beyond the evaluation of clinical care and into the realm of human rights and morality. Although few will argue that simulation is a convenient tool for competency assessment, many will question the transferability of this assessment in the simulation laboratory to the clinical environment. Two studies addressed this issue. In the first study, a group of emergency medicine residents at various levels of training were assessed on the basis of observable, objective events and time to completion of a surgical airway.⁹³ The simulation-based evaluation was able to discern novice residents from more advanced residents. A separate study compared anesthesiology residents at different levels of training with board-certified anesthesiologists.⁹⁴ These groups

were exposed to 12 unique, simulated intraoperative emergencies. Again, the simulation-based assessment distinguished the junior trainees from more experienced trainees and anesthesiologists.

Assessment using medical simulation can also be applied to evaluating impaired clinicians for continued practice or promotion. One article outlines the use of a human patient simulator in evaluating an anesthesiologist identified as having substandard medical skills.⁹⁵ In this report, the candidate underwent a psychological inventory, oral examinations, and a written test. These were performed to assess medical knowledge and suitability for remediation. The evaluation was enhanced by 4 simulated cases on a human patient simulator that was relevant to the candidate's practice. The simulation allowed an assessment of technical proficiency within the candidate's practice environment. In a separate report, some dentists with medical disabilities were evaluated in a simulated environment to determine their ability to perform clinically.⁹⁶

CONTINUING MEDICAL EDUCATION

This past century has witnessed an explosion in medical knowledge. With the ever-changing landscape of medicine, there must be a mechanism by which clinicians can stay current in their medical knowledge and practice. Most state medical licensing boards and specialty boards require some form of CME. Over the years, CME programs have evolved into many forms. Despite its wide acceptance as a requirement for ongoing medical practice, there is little evidence that participation in CME improves patient care or outcomes.⁹⁷ However, several trends exist regarding CME activities. Marinopoulos and colleagues⁹⁸ performed an exhaustive review regarding the effectiveness of CME. They found 3 common themes: live media are more effective than print media, interventions with multiple media are more effective than interventions with single media, and multiple exposures are more effective than a single exposure. Simulation-based CME programs fulfill the first 2 themes on this list. Simulation-based CME is usually offered in combination with didactic lectures. Because of the cost and logistics of performing simulation, high-fidelity simulation-based CME is not as common as part-task training CME exercises. One study describes the use of high-fidelity simulation for crisis resource management in anesthesia.⁹⁹ Blum and colleagues⁹⁹ proposed that simulation-based courses should be more widely used. A separate study examining the use

of a simulation-based course for difficult airway training suggests that this type of program is more conducive to affecting clinical practice.¹⁰⁰ Offering simulation-based CME requires a fair amount of resources. Simulators are costly, and a greater amount of time is necessary to perform simulations as opposed to lectures. Therefore, having sufficient faculty familiar with performing simulations can also be a problem. Several professional medical societies have recently begun accrediting simulation-based educational centers.^{101–104} Although each group has its own criteria for accreditation, they set standards for necessary equipment, space, and faculty. Theoretically, these centers would be adequately equipped to offer CME programs in their respective specialties.

FUTURE OF SIMULATION IN MEDICAL EDUCATION

Board Certification and Credentialing

The use of simulation technologies for undergraduate and graduate medical assessment has become well established in multiple specialties. However, use of these technologies for credentialing and certification is in its infancy. The United States Medical Licensing Examination (USMLE) was the first testing organization to integrate computer-based case simulations.¹⁰⁵ Examinees are provided a chief complaint and a brief history and are required to care for the virtual patient. These simulations include changes in the patient's status requiring multiple interventions. In 2004, the USMLE also added a clinical skills assessment using standardized patients. Examinees are required to participate in 12 encounters while the standardized patient records progress on a case-specific dichotomous checklist. Several specialty boards incorporate oral case presentations as part of the board certification process. The American Board of Family Practice expanded its board certification examination in family practice to include simulation.¹⁰⁶ This examination includes computer-based case simulations that are uniquely generated from a knowledge-base database for each examinee. The system uses population distributions for disease states to create patients that evolve in response to candidate interventions, such as pharmacological and nonpharmacological therapies. A third-party application uses Bayesian theory to analyze the participant's progress and manipulate the clinical case. This is a significantly more advanced system than the prefabricated cases on the USMLE examination. The Royal College

of Physicians and Surgeons of Canada (RCPSC) has long incorporated standardized patients into its comprehensive objective examination in internal medicine. However, the RCPSC quickly recognized the limitations of standardized patients in producing accurate physical examination findings. In 2003, it augmented its examination with digitized cardiac auscultation videos.¹⁰⁷ By including simulation, the RCPSC was able to include an element of high-stakes assessment that is not otherwise possible with standardized patients alone. The Israeli Board of Anesthesiology Examination Committee fully integrated several simulation platforms into its board certification examination.¹⁰⁸ The Israeli Board of Anesthesiology Examination Committee integrated a simulation-based objective structured clinical examination component into the board examination process. It created 5 standardized scenarios to assess trauma management, ACLS, operating room crisis management, mechanical ventilation, and administration of regional anesthesia. A combination of high-fidelity mannequin simulators, standardized patients, and an artificial lung was used in the testing environment. To date, the Israeli Board of Anesthesiology examination is the first to incorporate such a robust use of simulation technologies. Its results yielded only adequate interrater correlation, and thus further investigation is needed to improve this model of examination. As new evaluation tools are tested and validated, the various medical specialties may be more likely to adopt the use of simulation for board certification and credentialing. Simulation offers reproducibility and realism.

Medical-Legal Applications

The use of simulation also has a growing impact in the medical-legal arena. The Consolidated Risk Insurance Company (CRICO) is the patient safety and medical malpractice company owned by and serving the Harvard medical community. In 2001, CRICO began offering insurance premium incentives for anesthesiologists who participated in simulation-based crisis resource management.^{109,110} Upon analyzing its malpractice claims after several years of implementing this incentive, CRICO concluded that the program was effective in reducing the cost of malpractice claims. The benefit was so significant that CRICO implemented a similar program in obstetrics and gynecology.^{70,110} It is conceivable that other insurance companies will make this type of training required for other specialties. In the book *Practical Health Care Simulations*,¹¹ Eason provides a theoretical rendering of using simulation in the courtroom.

He discusses using medical simulation as evidence in a medical malpractice case. In his example, the medical case is recreated on a simulator for the jury. The intent is to prove or refute wrong doing by the physician. Although there are several legal considerations in using simulations as evidence, this is potentially a very interesting application of medical simulation.

CONCLUSION

Simulation in medical training is here to stay. From the learners' perspective, simulation affords the ideal opportunity to practice patient care away from the bedside and to apply the principles of adult learning and the principles of deliberate practice toward knowledge and skills mastery. From the patients' perspective, having students and residents pretrain with simulation increases the likelihood of a minimum competency level prior to clinical interaction and medical decision making. Patients are more willing to allow students to perform procedures on them after they have undergone simulation training.⁴⁷

There are various practical limitations to implementing simulation-training programs. One recent study, which surveyed all emergency medicine residencies using simulation, found faculty time constraints to be the top barrier to simulation use followed by a lack of faculty training and cost of equipment.³ Many of the technologies implemented by simulation educators are expensive. Depending on the fidelity of a simulator, the price for equipment alone can range from \$6000 to \$250,000. When this is coupled with the cost of equipment maintenance, space for educational labs, and personnel, the budget for a simulation center can be quite large. Many of these expenses can be minimized by the development of multidisciplinary centers in which funding is provided across various departments in an institution. Identifying qualified faculty skilled in simulation use and debriefing is another significant barrier. This includes availability in terms of protected time as well as those faculty trained to teach. This technology does not obviate the need for faculty trained in solid educational principles and teaching techniques. In other words, simulators do not replace good educators.

As the field of medical simulation continues to grow, an increasing number of educators are being formally trained to employ simulation. Currently, 4 emergency medicine programs offer fellowship training in simulation after residency.^{3,112} Ultimately, to increase the availability of simulation

for students and residents, institutions must prioritize simulation in their hospitals and medical schools by properly rewarding educators through promotions and compensating them with protected time to teach and do further research in this area.

The field of medical simulation continues to grow. In a past publication,¹¹³ Gaba predicted that simulation would be "embedded in the fabric of care" by the beginning of the 21st century. Advanced Initiatives in Medical Simulation (AIMS) was established to accomplish just that. AIMS is a coalition of individuals and organizations committed to promoting the use of medical simulation to improve patient safety, reduce medical error, and lower healthcare costs. With the efforts of AIMS, Rep. Randy Forbes (R-VA) and Rep. Patrick Kennedy (D-RI) introduced bill HR 4321, the Enhancing SIMULATION (Safety in Medicine Utilizing Leading Advanced Simulation Technologies to Improve Outcomes Now) Act of 2007.³¹ This bill would advance the use of medical simulation in research and healthcare training and would create medical simulation centers of excellence. Indeed, it is very likely that the future will bring the widespread use of simulation in all aspects of our healthcare infrastructure.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Gaba DM. The future vision of simulation in health care. *Qual Saf Health Care* 2004; 13(suppl 1): i2–i10.
- Berkenstadt H, Ziv A, Gafni N, et al. The validation process of incorporating simulation-based accreditation into the anesthesiology Israeli national board exams. *Isr Med Assoc J* 2006; 8: 728–733.
- Okuda Y, Bond WF, Bonfante G, et al. National growth in simulation training within emergency medicine residency programs, 2003–2008. *Acad Emerg Med* 2008; 15: 1113–1116.
- Gaba DM. Improving anesthesiologists' performance by simulating reality. *Anesthesiology* 1992; 76: 491–494.
- Kohn L, Corrigan J, Donaldson MS. *To Err Is Human: Building a Safer Health Care System*. Washington, DC: National Academy Press; 2000.
- Okuda Y, Quinones J. The use of simulation in the education of emergency care providers for cardiac emergencies. *Int J Emerg Med* 2008; 1: 73–77.
- Levine AI, Swartz MH. Standardized patients: the "other" simulation. *J Crit Care* 2008; 23: 179–184.
- Grenvik A, Schaefer J. From Resusci-Anne to SimMan: the evolution of simulators in medicine. *Crit Care Med* 2004; 32(suppl): S56–S57.
- Abrahamson S, Denson JS, Wolf RM. Effectiveness of a simulator in training anesthesiology residents. *J Med Educ* 1969; 44: 515–519.
- Cooper JB, Taqueti VR. A brief history of the development of mannequin simulators for clinical education and training. *Qual Saf Health Care* 2004; 13(suppl 1): i11–i18.
- Marsch S. Team oriented medical simulation. In: Hensen LC, Lee AC, eds. *Simulators in Anesthesiology Education*. New York, NY: Plenum Press; 1998; 51–55.
- Bryson EO, Levine AI. The simulation theater: a theoretical discussion of concepts and constructs that enhance learning. *J Crit Care* 2008; 23: 185–187.
- Bond WF, Deitrick LM, Arnold DC, et al. Using simulation to instruct emergency medicine residents in cognitive forcing strategies. *Acad Med* 2004; 79: 438–446.
- Bryan R, Kreuter M, Brownson R. Integrating adult learning principles into training for public health practice. *Health Promot Pract*. doi://10.1177/1524839907308117. Available at: <http://hpp.sagepub.com>.
- Misch DA. Andragogy and medical education: are medical students internally motivated to learn? *Adv Health Sci Educ Theory Pract* 2002; 7: 153–160.
- Croley WC, Rothenberg DM. Education of trainees in the intensive care unit. *Crit Care Med* 2007; 35(suppl): S117–S121.
- Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004; 79(suppl): S70–S81.
- Eppich WJ, Adler MD, McGaghie WC. Emergency and critical care pediatrics: use of medical simulation for training in acute pediatric emergencies. *Curr Opin Pediatr* 2006; 18: 266–271.
- Wayne DB, Butter J, Siddall VJ, et al. Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. *J Gen Intern Med* 2006; 21: 251–256.
- Wayne DB, Siddall VJ, Butter J, et al. A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Acad Med* 2006; 81(suppl): S9–S12.
- Issenberg SB, McGaghie WC, Petrusa ER, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005; 27: 10–28.
- McGaghie WC, Issenberg SB, Petrusa ER, et al. Effect of practice on standardised learning outcomes in simulation-based medical education. *Med Educ* 2006; 40: 792–797.
- Rolfe IE, Sanson-Fisher RW. Translating learning principles into practice: a new strategy for learning clinical skills. *Med Educ* 2002; 36: 345–352.
- Ziv A, Wolpe PR, Small SD, et al. Simulation-based medical education: an ethical imperative. *Acad Med* 2003; 78: 783–788.
- Chakraborti C, Boonyasai RT, Wright SM, et al. A systematic review of teamwork training interventions in medical student and resident education. *J Gen Intern Med* 2008; 23: 846–853.
- Ottstad E, Boulet JR, Lighthall GK. Evaluating the management of septic shock using patient simulation. *Crit Care Med* 2007; 35: 769–775.

27. Wayne DB, Didwania A, Feinglass J, et al. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. *Chest* 2008; 133: 56–61.
28. Savoldelli GL, Naik VN, Park J, et al. Value of debriefing during simulated crisis management: oral versus video-assisted oral feedback. *Anesthesiology* 2006; 105: 279–285.
29. Young JS, Dubose JE, Hedrick TL, et al. The use of “war games” to evaluate performance of students and residents in basic clinical scenarios: a disturbing analysis. *J Trauma* 2007; 63: 556–564.
30. Steadman RH, Coates WC, Huang YM, et al. Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. *Crit Care Med* 2006; 34: 151–157.
31. Kelsey R, Botello M, Millard B, et al. An online heart simulator for augmenting first-year medical and dental education. *Proc AMIA Symp* 2002; 370–374.
32. Via DK, Kyle RR, Trask JD, et al. Using high-fidelity patient simulation and an advanced distance education network to teach pharmacology to second-year medical students. *J Clin Anesth* 2004; 16: 144–151.
33. Seybert AL, Kobulinsky LR, McKaveney TP. Human patient simulation in a pharmacotherapy course. *Am J Pharm Educ* 2008; 72: 37.
34. Fitch MT. Using high-fidelity emergency simulation with large groups of preclinical medical students in a basic science course. *Med Teach* 2007; 29: 261–263.
35. Ewy GA, Felner JM, Juul D, et al. Test of a cardiology patient simulator with students in fourth-year electives. *J Med Educ* 1987; 62: 738–743.
36. Issenberg SB, Gordon MS, Greber AA. Bedside cardiology skills training for the osteopathic internist using simulation technology. *J Am Osteopath Assoc* 2003; 103: 603–607.
37. Issenberg SB, McGaghie WC, Gordon DL, et al. Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. *Teach Learn Med* 2002; 14: 223–228.
38. McIvor WR. Experience with medical student simulation education. *Crit Care Med* 2004 32(suppl): S66–S69.
39. Murray D, Boulet J, Ziv A, et al. An acute care skills evaluation for graduating medical students: a pilot study using clinical simulation. *Med Educ* 2002; 36: 833–841.
40. Boulet JR, Murray D, Kras J, et al. Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. *Anesthesiology* 2003; 99: 1270–1280.
41. Weller J, Robinson B, Larsen P, et al. Simulation-based training to improve acute care skills in medical undergraduates. *N Z Med J* 2004; 117: U1119.
42. Morgan PJ, Cleave-Hogg D, Desousa S, et al. Applying theory to practice in undergraduate education using high fidelity simulation. *Med Teach* 2006; 28: e10–e15.
43. Lucas S, Tuncel A, Bensalah K, et al. Virtual reality training improves simulated laparoscopic surgery performance in laparoscopy naive medical students. *J Endourol* 2008; 22: 1047–1051.
44. Van Sickle KR, Ritter EM, Smith CD. The pretrained novice: using simulation-based training to improve learning in the operating room. *Surg Innov* 2006; 13: 198–204.
45. Sanchez LD, Delapena J, Kelly SP, et al. Procedure lab used to improve confidence in the performance of rarely performed procedures. *Eur J Emerg Med* 2006; 13: 29–31.
46. Gardner R, Raemer DB. Simulation in obstetrics and gynecology. *Obstet Gynecol Clin North Am* 2008; 35: 97–127.
47. Graber MA, Wyatt C, Kasperek L, et al. Does simulator training for medical students change patient opinions and attitudes toward medical student procedures in the emergency department? *Acad Emerg Med* 2005; 12: 635–639, ix.
48. Morgan PJ, Cleave-Hogg D. A worldwide survey of the use of simulation in anesthesia. *Can J Anaesth* 2002; 49: 659–662.
49. Gaba DM, DeAnda A. The response of anesthesia trainees to simulated critical incidents. *Anesth Analg* 1989; 68: 444–451.
50. DeAnda A, Gaba DM. Unplanned incidents during comprehensive anesthesia simulation. *Anesth Analg* 1990; 71: 77–82.
51. Schwid HA, O'Donnell D. Anesthesiologists' management of simulated critical incidents. *Anesthesiology* 1992; 76: 495–501.
52. Chopra V, Gesink B, de Jong J, et al. Does training on an anaesthesia simulator lead to improvement in performance? *Br J Anaesth* 1994; 73: 293–297.
53. Schwid HA, Rooke GA, Michalowski P, et al. Screen-based anesthesia simulation with debriefing improves performance in a mannequin-based anesthesia simulator. *Teach Learn Med* 2001; 13: 92–96.
54. Nishisaki A, Keren R, Nadkarni V. Does simulation improve patient safety? Self-efficacy, competence, operational performance, and patient safety. *Anesthesiol Clin* 2007; 25: 225–236.
55. American Board of Anesthesiology. Maintenance of certification in anesthesiology. Available at: <http://www.theaba.org/anesthesiology-maintenance.shtml#components>. Accessed August 2008.
56. Ziv A, Rubin O, Sidi A, et al. Credentialing and certifying with simulation. *Anesthesiol Clin* 2007; 25: 261–269.
57. Chatenay M, Maguire T, Skakun E, et al. Does volume of clinical experience affect performance of clinical clerks on surgery exit examinations? *Am J Surg* 1996; 172: 366–372.
58. Sachdeva AK, Russell TR. Safe introduction of new procedures and emerging technologies in surgery: education, credentialing, and privileging. *Surg Clin North Am* 2007; 87: 853–866, vi–vii.
59. Sachdeva AK. Invited commentary: educational interventions to address the core competencies in surgery. *Surgery* 2004; 135: 43–47.
60. Sachdeva AK. Acquiring skills in new procedures and technology: the challenge and the opportunity. *Arch Surg* 2005; 140: 387–389.
61. Kneebone R. Simulation in surgical training: educational issues and practical implications. *Med Educ* 2003; 37: 267–277.
62. Crofts JF, Fox R, Ellis D, et al. Observations from 450 shoulder dystocia simulations: lessons for skills training. *Obstet Gynecol* 2008; 112: 906–912.

63. Halvorsen FH, Elle OJ, Fosse E. Simulators in surgery. *Minim Invasive Ther Allied Technol* 2005; 14: 214–223.
64. Boyd KB, Olivier J, Salameh JR. Surgical residents' perception of simulation training. *Am Surg* 2006; 72: 521–524.
65. Dunkin B, Adrales GL, Apelgren K, et al. Surgical simulation: a current review. *Surg Endosc* 2007; 21: 357–366.
66. Sedlack RE, Baron TH, Downing SM, et al. Validation of a colonoscopy simulation model for skills assessment. *Am J Gastroenterol* 2007; 102: 64–74.
67. Aggarwal R, Ward J, Balasundaram I, et al. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg* 2007; 246: 771–779.
68. Stefanidis D, Korndorffer JR Jr, Sierra R, et al. Skill retention following proficiency-based laparoscopic simulator training. *Surgery* 2005; 138: 165–170.
69. Korndorffer JR Jr, Dunne JB, Sierra R, et al. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *J Am Coll Surg* 2005; 201: 23–29.
70. Gardner R, Walzer TB, Simon R, et al. Obstetric simulation as a risk control strategy: course design and evaluation. *Simul Healthc* 2008; 3: 119–127.
71. Maher JE, Kleinman GE, Lile W, et al. The construction and utility of an amniocentesis trainer. *Am J Obstet Gynecol* 1998; 179: 1225–1227.
72. Dupuis O, Moreau R, Silveira R, et al. A new obstetric forceps for the training of junior doctors: a comparison of the spatial dispersion of forceps blade trajectories between junior and senior obstetricians. *Am J Obstet Gynecol* 2006; 194: 1524–1531.
73. Deering S, Poggi S, Macedonia C, et al. Improving resident competency in the management of shoulder dystocia with simulation training. *Obstet Gynecol* 2004; 103: 1224–1228.
74. Crofts JF, Bartlett C, Ellis D, et al. Management of shoulder dystocia: skill retention 6 and 12 months after training. *Obstet Gynecol* 2007; 110: 1069–1074.
75. Draycott TJ, Crofts JF, Ash JP, et al. Improving neonatal outcome through practical shoulder dystocia training. *Obstet Gynecol* 2008; 112: 14–20.
76. Crofts JF, Bartlett C, Ellis D, et al. Documentation of simulated shoulder dystocia: accurate and complete? *BJOG* 2008; 115: 1303–1308.
77. Crofts JF, Bartlett C, Ellis D, et al. Training for shoulder dystocia: a trial of simulation using low-fidelity and high-fidelity mannequins. *Obstet Gynecol* 2006; 108: 1477–1485.
78. Goffman D, Heo H, Pardani S, et al. Improving shoulder dystocia management among resident and attending physicians using simulations. *Am J Obstet Gynecol* 2008; 199: 294.
79. American College of Obstetricians and Gynecologists. Post partum hemorrhage. ACOG practice bulletin. *Obstet Gynecol* 2006; 108: 1039–1047.
80. Maslovitz S, Barkai G, Lessing JB, et al. Recurrent obstetric management mistakes identified by simulation. *Obstet Gynecol* 2007; 109: 1295–1300.
81. DeVita MA, Schaefer J, Lutz J, et al. Improving medical emergency team (MET) performance using a novel curriculum and a computerized human patient simulator. *Qual Saf Health Care* 2005; 14: 326–331.
82. Reznick M, Smith-Coggins R, Howard S, et al. Emergency medicine crisis resource management (EMCRM): pilot study of a simulation-based crisis management course for emergency medicine. *Acad Emerg Med* 2003; 10: 386–389.
83. Fritz PZ, Gray T, Flanagan B. Review of mannequin-based high-fidelity simulation in emergency medicine. *Emerg Med Australas* 2008; 20: 1–9.
84. Carbine DN, Finer NN, Knodel E, et al. Video recording as a means of evaluating neonatal resuscitation performance. *Pediatrics* 2000; 106: 654–658.
85. Thomas EJ, Sexton JB, Lasky RE, et al. Teamwork and quality during neonatal care in the delivery room. *J Perinatol* 2006; 26: 163–169.
86. Halamek LP, Kaegi DM, Gaba DM, et al. Time for a new paradigm in pediatric medical education: teaching neonatal resuscitation in a simulated delivery room environment. *Pediatrics* 2000; 106: E45.
87. Hunt EA, Shilkofski NA, Stavroudis TA, et al. Simulation: translation to improved team performance. *Anesthesiol Clin* 2007; 25: 301–319.
88. Britt RC, Reed SF, Britt LD. Central line simulation: a new training algorithm. *Am Surg* 2007; 73: 680–682; discussion 682–683.
89. Dibb C. Medical residency: when are program administrators liable? *J Legal Educ* 2007; 281: 1–8.
90. Bond WF, Spillane L. The use of simulation for emergency medicine resident assessment. *Acad Emerg Med* 2002; 9: 1295–1299.
91. Wang EE, Vozenilek JA. Addressing the systems-based practice core competency: a simulation-based curriculum. *Acad Emerg Med* 2005; 12: 1191–1194.
92. Gisondi MA, Smith-Coggins R, Harter PM, et al. Assessment of resident professionalism using high-fidelity simulation of ethical dilemmas. *Acad Emerg Med* 2004; 11: 931–937.
93. Girzadas DV Jr, Clay L, Caris J, et al. High fidelity simulation can discriminate between novice and experienced residents when assessing competency in patient care. *Med Teach* 2007; 29: 452–456.
94. Murray DJ, Boulet JR, Avidan M, et al. Performance of residents and anesthesiologists in a simulation-based skill assessment. *Anesthesiology* 2007; 107: 705–713.
95. Rosenblatt MA, Abrams KJ. The use of a human patient simulator in the evaluation of and development of a remedial prescription for an anesthesiologist with lapsed medical skills. *Anesth Analg* 2002; 94: 149–153.
96. Raborn GW, Carter RM. Using simulation to evaluate clinical competence after impairment. *J Can Dent Assoc* 1999; 65: 384–386.
97. Schrock JW, Cydulka RK. Lifelong learning. *Emerg Med Clin North Am* 2006; 24: 785–795.
98. Marinopoulos SS, Dorman T, Ratanawongsa N, et al. Effectiveness of continuing medical education. *Evid Rep Technol Assess (Full Rep)* 2007; 149: 1–69.
99. Blum RH, Raemer DB, Carroll JS, et al. Crisis resource management training for an anaesthesia faculty: a new approach to continuing education. *Med Educ* 2004; 38: 45–55.
100. Russo SG, Eich C, Barwing J, et al. Self-reported changes in attitude and behavior after attending a simulation-aided airway management course. *J Clin Anesth* 2007; 19: 517–522.
101. Sachdeva AK, Pellegrini CA, Johnson KA. Support for simulation-based surgical education through American College of Surgeons-accredited education institutes. *World J Surg* 2008; 32: 196–207.

102. Sinz EH. Anesthesiology national CME program and ASA activities in simulation. *Anesthesiol Clin* 2007; 25: 209–223.
103. American Society of Anesthesiologists. ASA approval of anesthesiology simulation programs. Available at: <http://www.asahq.org/sim/whitepaper.htm>. Accessed August 2008.
104. Society for Simulation in Healthcare. SSH accreditation standards for simulation programs in healthcare (draft, June 2008). Available at: http://www.ssih.org/public/index.php?ijejifeijfjfwf=qhuihqwiduhqwiufe&page=website_news. Accessed August 2008.
105. Dillon GF, Boulet JR, Hawkins RE, et al. Simulations in the United States Medical Licensing Examination (USMLE). *Qual Saf Health Care* 2004; 13(suppl 1): i41–i45.
106. Hagen MD, Sumner W, Roussel G, et al. Computer-based testing in family practice certification and recertification. *J Am Board Fam Pract* 2003; 16: 227–232.
107. Hatala R, Kassen BO, Nishikawa J, et al. Incorporating simulation technology in a Canadian internal medicine specialty examination: a descriptive report. *Acad Med* 2005; 80: 554–556.
108. Berkenstadt H, Ziv A, Gafni N, et al. Incorporating simulation-based objective structured clinical examination into the Israeli national board examination in anesthesiology. *Anesth Analg* 2006; 102: 853–858.
109. McCarthy J, Cooper JB. Malpractice insurance carrier provides premium incentive for simulation-based training and believes it has made a difference. *Anesth Patient Saf Found News* 2007; 22: 17.
110. Hanscom R. Medical simulation from an insurer's perspective. *Acad Emerg Med* 2008; 15: 984–987.
111. Eason M. Community and legal services. In: Loyd G, Lake C, Greenberg R, eds. *Practical Health Care Simulations*. Philadelphia, PA: Elsevier; 2004: 565–571.
112. McLaughlin S, Fitch MT, Goyal DG, et al. Simulation in graduate medical education 2008: a review for emergency medicine. *Acad Emerg Med* 2008; 15: 1117–1129.
113. Enhancing SIMULATION (Safety in Medicine Utilizing Leading Advanced Simulation Technologies to Improve Outcomes Now) Act of 2007, HR 4321, 110th Cong, 1st Sess 2007.

Student-Run Health Clinic: Novel Arena to Educate Medical Students on Systems-Based Practice

Yasmin S. Meah, MD, Eric L. Smith, PhD, and David C. Thomas, MD

Mount Sinai School of Medicine, New York, NY

ABSTRACT

In recent decades, the United States has experienced substantial growth in the number of student-run clinics for the indigent. Today, over 49 medical schools across the country operate over 110 student-run outreach clinics that provide primary care services to the poor and uninsured. Despite this development, little research has been published on the educational value of such student-led endeavors. Although much has been surmised, no general methodology for categorizing the learning experience in these clinics has been established. This article represents the first literature review of the novel method of educating students through the operation of a clinic for the underserved. It highlights the student-run clinic as a unique enhancement of medical education that may supplant current curricular arenas in teaching students about systems-based practice principles such as cost containment and financing, resource allocation, interdisciplinary collaboration, patient advocacy, and monitoring and delivery of quality care. The novelty of the student-run clinic is that students place themselves at the forefront of problem solving and system navigation to effectively care for severely disadvantaged populations. This article underscores the student-run clinic as a potentially ideal experiential learning method for preparing young physicians to confront a US healthcare system currently facing crises in cost, quality of care, and high rates of uninsurance. The article stresses the need for outcomes research on the long-term effectiveness of the student-run clinic experience in affecting medical student practice behaviors and attitudes in patient care settings that extend beyond

the student-run clinic. *Mt Sinai J Med* 76:344–356, 2009. © 2009 Mount Sinai School of Medicine

Key Words: ACGME core competencies, experiential learning, student-run clinic, systems-based practice, uninsured.

It is a well-accepted phenomenon that medical students of this generation are inheriting a healthcare system in crisis.¹ Recent Institute of Medicine reports have highlighted the inability of the current healthcare delivery system to meet the present and future needs of the US public.^{2–4} Quality and cost of medical care are and will continue to remain the predominant stressors of the current US healthcare system.^{5,6} Both directly affect the access to care for disadvantaged populations and necessitate dramatic changes in the current organization and delivery of healthcare to make it fair and unbiased. In response, the seminal 2001 report by the Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*, urged medical training programs to change the way in which physicians are educated to meet the “six aims for improvement” and the “ten rules for redesign of the healthcare system” necessary to fundamentally improve the equitability of healthcare for US residents.² The Accreditation Council on Graduate Medical Education (ACGME) has responded similarly, requiring residency programs to implement substantial changes to achieve 6 core competency goals that all residents should meet.⁷ Systems-based practice, one of the most difficult to implement,⁸ is the goal that underscores the stress of the ACGME on the need for rising physicians to comprehensively understand and navigate the system in which they will practice to provide the most optimal care to the diverse population of patients that they will service.⁶ Despite this evolution in healthcare training for residents in response to the increasing pressure by accrediting and expert committees to address

Address Correspondence to:

Yasmin S. Meah

Mount Sinai School of Medicine
New York, NY

Email: yasmin.meah@mssm.edu

this healthcare climate, curricular change at the undergraduate level has lagged.

In a 2005 national survey, Agrawal and colleagues¹ found that nearly half of their medical school respondents were dissatisfied with their current coursework in addressing health policy, healthcare delivery, and healthcare reform. Nearly 90% of their respondents wanted increased exposure to these essential issues. Medical students have thus joined ranks with policymakers, medical educators, and physicians in their serious concern that they are not being adequately prepared to deliver optimal healthcare in a system in which they will ultimately practice.

The national Undergraduate Medical Education for the 21st Century (UME 21) project marks a governmental response to improve the education of medical students in acquiring the knowledge, skills, and attitudes to practice in this healthcare system.^{5,9} Eighteen schools participated in this project through a competitive selection process; 8 of these schools undertook comprehensive roles over 3 years. The innovative part of this project was the collaboration and incorporation of external healthcare organizations including managed healthcare organizations and other community-based settings outside the traditional teaching hospital. The focus was not on managed care models but rather on the management of care with the realization that the environments of practice for graduates would continually evolve. Rivo and colleagues⁶ described various methods to teach systems-based care at the 8 partner UME 21 schools. Despite the wide variability in approaches among the schools, the authors concluded that a "comprehensive clinical experience beyond the traditional academic medical center teaching hospital where only about 1 in 1000 patients end up seeking medical care" was a necessary ingredient. Experiential learning in the context of direct patient care was a key element to a successful paradigm. At 1 UME 21 partner site, a 1-day educational program composed of a structured experience in addition to focused didactics positively changed medical students' attitudes toward and knowledge of healthcare delivery systems significantly more than didactics alone.⁹ At another UME 21 site, a longitudinal 6-month clinical experience in which students partnered with physicians and other healthcare providers in the provision of healthcare to a panel of patients significantly increased student understanding of the knowledge, skills, and responsibilities of physicians working in integrated delivery systems.⁹ The UME 21 project evaluators concluded that learners were more likely to retain systems-based core concepts in the ambulatory care clinical arena, where direct hands-on navigation of systems and

understanding of relationships between the various providers and provisioners of healthcare occurred.^{6,9}

Over the past decade, the United States has experienced a substantial proliferation of student-run clinics (SRCs) for the underserved, which supplement an existing potent yet overburdened safety net of privately operated, publicly operated, and volunteer-operated health clinics.¹⁰⁻¹² The first SRCs arose in the mid-1960s in various cities across the country, all as student-initiated endeavors with a commitment to the underserved.^{13,14} The current models for these SRCs vary widely; however, the standout theme is a healthcare delivery program for predominantly uninsured persons in which undergraduate and medical students manage with supervision the administrative aspects of all components of care.¹¹ Simpson and Long¹² were the first and only recent authors to survey and describe on a national scale the prevalence and operation of medical student-run health clinics. The authors were successful in retrieving data about the operations, finances, patient demographics, and acuity of complaints for 59 of 111 SRCs, the latter representing 49 medical schools across the country. The authors concluded, however, that despite these data, they could make no concrete assessment of the quality of clinical education in this arena.

Indeed, the data on this is far from robust in the literature. Nonetheless, from the 20 articles to date in the literature that address SRCs, we can draw meaningful conclusions about the educational value of such endeavors that may inform medical schools on the need to support such projects. In this article, we argue that the SRC may be an excellent experiential learning method for preparing young physicians to confront a US healthcare system currently facing crises in cost, quality of care, and high rates of uninsurance.

CURRENT LITERATURE ON STUDENT-RUN CLINICS

The existing literature on SRCs is in the early stages of evolution and thus is quite limited. Nonetheless, there are several promising studies that have emerged in recent years, emphasizing the SRC as an incubator of meaningful study and research that can both enrich the education of students and have broad-ranging implications for clinics that service the uninsured and other disadvantaged populations.

Table 1 depicts the current body of literature on SRCs organized by year of publication. In the mid-1990s, a number of descriptive communications

Table 1. *Current Literature on Student-Run Clinics Sorted by Year of Publication.*

Year of Publication	Authors	Sponsoring School of Medicine	Title of Article	Format of Article Study
1985	Campos-Outcalt ¹⁵	University of California Davis	Specialties Chosen by Medical Students Who Participated in a Student-Run, Community-Based Free Clinic	Observational retrospective cohort study of volunteers and career choices
1995	Cohen ¹⁶	Hahnemann	Eight Steps for Starting a Student-Run Clinic	Prescriptive short communication on program operation
1995	Collins ¹⁷	Hahnemann	The Hahnemann Homeless Clinics Project: Taking Health Care to the Streets and Shelters	Descriptive short communication on program operation
1995	Pi ¹⁸	University of California Davis	The Asian Clinic at UC Davis: Serving a Minority Population for Two Decades	Descriptive short communication on program operation
1995	Poulsen ¹¹	None	Student-Run Clinics: A Double Opportunity	Editorial
1995	Yap and Thornton ⁸	Stanford	The Arbor Free Clinic at Stanford: A Multidisciplinary Effort	Descriptive short communication on program operation
1996	Haq <i>et al.</i> ¹⁹	University of Wisconsin-Madison	Student and Faculty Collaboration in a Clinic for the Medically Underserved	Descriptive program development, implementation, and operation
2000	Davenport ²⁰	University of California San Francisco	Witnessing and the Medical Gaze: How Medical Students Learn to See at a Free Clinic for the Homeless	Descriptive bioethics of an student-run clinic
2001	Steinbach <i>et al.</i> ²¹	University of California Berkeley	The Berkeley Suitcase Clinic: Homeless Services by Undergraduate and Medical Student Teams	Descriptive program development, implementation, and operation
2003	Clark <i>et al.</i> ²²	Baylor	A Multidisciplinary, Learner-Centered, Student-Run Clinic for the Homeless	Descriptive program development, implementation, and operation
2004	Bennard <i>et al.</i> ²⁴	East Tennessee State University	A Student-Run Outreach Clinic for Rural Communities in Appalachia	Descriptive program development, implementation, and operation
2005	Beck ²³	University of California San Diego	The UCSD Student-Run Free Clinic Project: Transdisciplinary Health Professional Education	Descriptive program development, implementation, and operation
2005	Hurtado-Rogers <i>et al.</i> ²⁵	Mount Sinai School of Medicine	Education Through Service: A Medical Student-Facilitated Benefits Enrollment Initiative	Descriptive abstract-program development, implementation, and operation

(continued overleaf)

Table 1. (Continued).

Year of Publication	Authors	Sponsoring School of Medicine	Title of Article	Format of Article Study
2006	Buchanan and Witten ²⁶	Rush	Balancing Service and Education: Ethical Management of Student-Run Clinics	Descriptive bioethics of student run clinics
2006	Gilkey and Earp ¹³	University of North Carolina at Chapel Hill	Effective Interdisciplinary Training: Lessons from the University of North Carolina's Student Health Action Coalition	Observational cross-sectional study of clinic participants
2006	Moskowitz <i>et al.</i> ²⁷	University of Washington	Students in the Community: An Interprofessional Student-Run Free Clinic	Descriptive program development, implementation, and operation
2006	Niescierenko <i>et al.</i> ²⁸	University of Buffalo	Insuring the Uninsured: A Student-Run Initiative to Improve Access to Care in an Urban Community	Prospective and cross-sectional program evaluation
2007	Cadzow <i>et al.</i> ²⁹	University of Buffalo	The Health Status of Patients of a Student Run Free Medical Clinic in Inner-City Buffalo, NY	Observational cross-sectional study of clinic population characteristics
2007	Hastings <i>et al.</i> ³⁰	University of California Los Angeles	UCLA Mobile Clinic Project	Descriptive program development, implementation, and operation
2007	Simpson and Long ¹²	University of Pennsylvania	Medical Student-Run Health Clinics: Important Contributors to Patient Care and Medical Education	Observational cross-sectional study of national student-run clinics
Undated	Nordling ¹¹	None	Starting a Student-Run Homeless Clinic	Guide for program development, implementation, and operation
2008	Ryskina <i>et al.</i> ³¹	Mount Sinai School of Medicine	Quality of Diabetes Care at a Student-Run Community Health Clinic: How Does Free Clinic Care Compare to Publicly and Privately Insured Populations?	Abstract observational retrospective cohort study
2008	Jimenez <i>et al.</i> ³²	University of Medicine and Dentistry—Robert Wood Johnson	The Promise Clinic: A Service Learning Approach to Increasing Access to Health Care	Descriptive program development, implementation, and operation

emerged congruous with the burgeoning of SRCs across the country. Although over 110 clinics currently exist, only 14 articles currently address program development and logistics; the majority of these articles materialized after 2003. Since 1985, 5 observational studies, however, have also come forth, all authored primarily by medical students. Most notably, the article by Simpson and Long¹² of the University of Pennsylvania stands out as the most comprehensive, identifying program characteristics for clinics across the United States. The 2 articles to emerge out of the University of Buffalo^{28,29} and the article authored by Gilkey and Earp¹³ of the University of North Carolina are the most methodologically rigorous SRC program evaluations currently in the literature. To date, the authors know of only 1 additional abstract authored by medical students at the Mount Sinai School of Medicine that uses systematic methodology in program evaluation through an observational retrospective study.³¹

CONCLUSIONS FROM THE LITERATURE ON STUDENT-RUN CLINICS: USING CONSTRUCTIVIST METHODS TO PREPARE STUDENTS TO INHERIT A HEALTHCARE SYSTEM IN CRISIS

For medical educators to more completely appreciate the method and value of learning in an SRC, it is useful to understand the SRC within the context of constructivist learning theory. Current trends in medical education value constructivism as an ideal approach to teaching students to conceptualize medical disease and care for patients.³³ Two key elements of cognitive constructivism are the promotion of student independence and active learning techniques.³⁴ To understand and champion constructivism is to appreciate the value of self-directed learning as the most effective approach to contemporary medical education.³⁵ Self-directed learning means quite literally the process in which students take the initiative for their own learning. In line with adult learning theory, self-directed learning is most cultivated in situations in which education is based on experience and when learning occurs through opportunities in which new knowledge and understanding can be integrated into the personal and professional context of the learner.³⁶

Experiential learning remains a fundamental element of medical education³⁷ and is a critical strategy to foster self-directed learning.³⁵ Traditionally, medical students in their third and fourth years are placed

on the hospital wards to experience and apply clinical knowledge to real-life scenarios in real time. As constructivism has taken hold in medical education, educators have appreciated the role of experiential learning methods in reinforcing abstract physiological and psychosocial concepts in the preclinical years.³⁸ Williams and Deci³⁹ more recently highlighted the importance of upholding student autonomy in medical education as critical to nurturing more responsible and humanistic learning and execution of patient care.

Unlike traditional clinical settings, the SRC offers one of the most concrete and comprehensive examples of experiential learning in which students operate more autonomously in directing their own learning than in other more traditional venues. As such, it offers the constructs of a novel educational paradigm that deserves to be examined with respect to its potentially influential and wide-ranging effects on medical student education.

How does the SRC differ from existing models of education in medical schools? Unlike traditional clinical settings in which students train, the SRC is staffed primarily by medical students who function largely as directors of daily and global operations of the clinic, navigating a number of complex administrative duties for which in large part they receive little to no training in traditional medical school settings.^{11,23} These functions differ from clinic to clinic but commonly include tasks and problem solving that relate to the following:

1. Clinic flow^{12,17,18,26,32};
 - a. Patient registration and scheduling.
 - b. Daily patient, student volunteer, ancillary staff, and physician work flow.
 - c. Delegation of student, physician, and interdisciplinary staff duties.
 - d. Daily and weekly follow-up of patient issues and turnover to ensure continuous patient care.
 - e. Clinic site organization, maintenance, and assurance of adequate supply stock.
2. Endeavor design, modification, and sustainability.^{12,23,24,32};
 - a. Development of and adherence to a mission statement.
 - b. Design and implementation of student leadership and annual turnover.
 - c. Resource acquisition (funds, supplies, specialty services, laboratories, personnel, and benefits) and cost-effective allocation to patients (eg, designing drug formularies to restrict patient-related costs).

- d. Successful incorporation of interdisciplinary personnel.^{13,23,25}
3. Outreach and community presence:
 - a. Creation of and adherence to a mission statement and alignment and realignment of activities to meet the goals of the community.³²
 - b. Creation and sustainment of community partnerships and outreach efforts.^{24,32}
4. Student education:
 - a. Oversight of student clinical teams (in many of these clinics, 1 or more junior students are paired with 1 or more senior students who as a team evaluate patients, make treatment decisions, formulate plans, and present cases for final evaluation and treatment intervention to site preceptors).^{23,24,32,40}
 - b. Direction of student educational missions of a clinic.
 - c. Oversight and direction of clinic-based research.²⁹
5. Quality evaluation and control:
 - a. Understanding, evaluating, and ensuring best practices.³¹
 - b. Understanding clinic population needs and continuously modifying clinic services and structure.^{29,32}

The restrictions on these SRCs make creative management and delivery of effective healthcare a constant challenge uniquely experienced outside of traditional medical arenas. In conventional terms, the clinics do not exist as traditional delivery systems of healthcare and thus pose distinctive problems that students may otherwise not encounter in medical school. For example, nearly 90% of these clinics primarily service the uninsured, and thus the majority take no payments for services, relying on fundraising, institutional support, and grants for operations.¹² Often, the students in leadership positions are the primary seekers of funds to support the clinic³²; thus, students must possess an intimate knowledge of program finances and an anticipation of costs to fund programs in evolution in order to successfully manage and develop the SRC.

Additionally, the operation and reliance of the SRC on volunteers, including physicians, nurses, and social workers, in addition to medical students restrict the days of operation. The majority of clinics surveyed by Simpson and Long¹² remained open only once a week; this differs dramatically from traditional delivery systems, which typically operate 5 to 7 days a week. Thus, the management of crises and issues that arise during the week as well as the accommodation of ill patients on

often overlooked clinic days requires students to skillfully triage patients and negotiate resources and manpower to effectively address patient needs.^{32,40} Furthermore, when these clinics are in session, a number of them are operating in unconventional sites that in some cases are improvised, temporary, and stocked with limited supplies that are otherwise more readily available in traditional clinic sites.¹² The locations vary widely, with the most eccentric being homeless shelters, warehouses, churches, and mobile units.^{17,18,26} In the majority of cases, however, the students have been successful in negotiating more equipped traditional office spaces,⁴⁰ holding clinics during off hours such as weekends and nights when staff personnel are absent; thus, administrative and ancillary clinical duties such as phlebotomy are predominantly or exclusively performed by volunteer students.

Creative management and delivery of adequate healthcare is thus a constant challenge for these students and can foster a deeper, more hands-on appreciation of the inequities of healthcare and strategies to overcome them.¹⁸ Staffing SRCs may empower these students to more comprehensively understand the current healthcare landscape, thus transforming these students into more effective current and future advocates for disadvantaged patients. On the other hand, the SRC may pose such unique challenges that students may find it difficult to extrapolate to traditional venues in which variations in insurance status are the norm. Unfortunately, no studies have examined how well the SRC prepares students for the challenges of navigating healthcare in more wide-ranging venues.

What the literature does emphasize, however, is the true uniqueness of the SRC in fostering the partnerships of students to lead, manage, and implement the overall health delivery system and educational mission of clinics that serve uninsured populations in their community. All of these clinics do so under the guidance of faculty, but the students involved in leadership positions far outnumber the physicians involved; thus, the responsibility of steering the healthcare of an SRC's population in large part falls on the student leadership.⁴⁰ Students often learn from one another through both administrative tasks and clinical tasks; this is in marked contrast to traditional delivery systems, in which administrative tasks are handled by personnel and networks outside the medical team: in many cases, these personnel and their varied tasks remain anonymous to students. Thus, students in traditional arenas may experience a more restricted sense of healthcare that is momentary and one-dimensional in contrast to students involved in an SRC, who may

experience a more comprehensive understanding of healthcare as executed through a collaboration of multiple personnel who operate chronically on a continuum of responsibility.

Arguably, the SRC harbors the potential for significant ethical implications in the training of medical students. These ethical debates are unique lessons in their own right and when positively construed can have lasting effects that promote advocacy among medical students involved in an SRC.^{20,40} In traditional delivery systems, medical students are not at the forefront of care; in the SRC, on the other hand, students are often navigating profoundly difficult tasks that they may have little to no experience in handling. In an article on the ethical controversy of students operating clinics, Buchanan and Witlen²⁶ identified the inherent moral conundrum as follows:

Students . . . have difficulty fulfilling these responsibilities due to their inexperience with clinical care, the healthcare system and the administrative structure of medical schools. Although students are legally bound not to practice medicine without the direct oversight of licensed physicians, their roles in organizational tasks that affect patient care are often not well defined.

The authors go on to identify a possible hidden curriculum in allowing such clinics to exist as an approval of students' practicing of skills on the impoverished and the acceptability of providing potentially subordinate care to the most disconnected from the healthcare system. A rigorous inspection of the literature, however, reveals no evidence that students in SRCs are in fact offering subordinate care or that they and their faculty leadership are propagating the impression that lesser care to the disenfranchised is acceptable. On the contrary, the literature supports the efforts of students to understand and equalize if not exceed traditional standards of care.^{28,29,31} Students who are in positions of managing an SRC may have a more problem-based understanding of the limitations of cost and resources, may have a more potent understanding of how to overcome barriers in order to equalize healthcare, and may thus be in more informed and decisive positions to navigate and direct a healthcare system in crisis when they ultimately embark on their careers. Under proper guidance and with mindful managing, SRCs can be an arena for socially conscious and responsible education where the best standards of care, in the face of limitations in resources, are understood, promulgated, and advocated for

all patients regardless of socioeconomic status and setting of care.¹¹

LIMITATIONS OF CONCLUSIONS FROM THE LITERATURE ON STUDENT-RUN CLINICS

The constricted literature on SRCs makes a number of vital conclusions nearly impossible. Drawing any meaningful supposition about just how many students are involved in these clinics is currently not feasible; thus, the magnitude of the potential impact of this learning modality is currently unclear. Furthermore, it is not possible to quantify and qualify the level of involvement of students working in these clinics. Invariably, some students sit on steering committees and make global decisions about the running of the clinic and its mission; others are only transiently involved. Whether the degree and level of clinic involvement positively translate into more substantial or lasting changes in student behaviors and the understanding of the US healthcare system is unclear. Alternatively, there are other venues at the medical school—voluntary and involuntary—that may comparatively translate into formidable changes in the understanding and navigation of a healthcare system that have not been measured by the existing literature on SRCs.

Finally, if supervision is fundamental to an ethically meaningful and morally sound educational experience in servicing such a vulnerable population in healthcare,¹¹ the existing literature currently fails to elaborate the extent and quality of faculty oversight of SRCs. To cement the validity of the SRC as a critical educational experience, the literature will need to illustrate qualitative and quantitative data on the degree of faculty guidance.

POTENTIAL LESSONS FROM STUDENT-RUN CLINICS: NEW PARADIGMS FOR LEARNING SYSTEMS-BASED PRACTICE

SRCs offer the essential backdrop for learning systems-based practice through a hands-on service-learning experience. Although in large part these clinics were originally envisioned and implemented as service projects to inculcate idealism and promote activism among students in collaboration with communities in need,^{8,17,18,23,41,42} the current literature has exposed these clinics as potentially critical

implementers of several core ACGME competencies. Effective systems-based practice will be highlighted here; this core competency demands future practitioners of healthcare to be able to (1) understand the global meaning of healthcare practices and the effect and influence of other healthcare professionals in the delivery of healthcare; (2) partner with other disciplines in a collegial manner that promotes effective assessment, coordination, and improvement of healthcare practices; (3) practice cost-effective healthcare and resource allocation that maintains the highest standards of quality care; (4) advocate for the highest standards of care for their patients; and (5) examine behaviors within one's practice and the larger context of system behaviors to identify and remediate singular and systematic errors and barriers in healthcare delivery.³ Although this is a core competency viewed as essential for graduating residents, the SRC in many cases harbors some of the earliest learning experiences in which such elements of systems-based practice can be taught.

CATEGORIZING SYSTEMS-BASED PRACTICE ELEMENTS IN THE STUDENT-RUN CLINIC

Interdisciplinary and Team-Based Care

Interdisciplinary care is a central facet of many of the existing SRCs.^{13,22,23,24,26,43} The University of California San Diego (UCSD) Student-Run Free Clinic Project highlights this as a key feature of its clinic, noting that one of the students' core mentors is a nurse practitioner.²³ Students at this site are supervised by representatives of various disciplines, including social workers, family therapists, and physicians, who collaborate in case management services that fall outside traditional medical services. Dental faculty share a central role in educating medical students about the social implications of poor oral hygiene, thus highlighting the more expansive nature of preventative health that takes place at an SRC.

The Houston Outreach Medicine, Education, and Social Services initiative of Baylor College of Medicine²² also highlights multidisciplinary care as an essential focus of the student-run endeavor. Students and faculty from the schools of public health, medicine, and pharmacy work together in teams to provide care to their predominantly homeless population. Participants ranked a multidisciplinary approach to care as one of the most valuable lessons gained from this experience.

The Berkeley Suitcase Clinic, in which public health students, medical students, and optometry students partner to deliver healthcare to the homeless, has underscored crucial differences between team care at a student-run free clinic and team care in more formal medical arenas.²¹ Feeling less pressured to perform, students at this SRC were more likely to identify and reach solutions for effective teamwork in a setting in which they felt more autonomous.

Bennard *et al.*²⁴ were the first to systematically qualify student responses to the educational value of working at the student-run outreach southern Appalachian clinics of East Tennessee State University. Over 75% responded that the experience positively affected their attitudes in working cooperatively with nonphysician health professionals.

The SRC provides an important context for understanding the effectiveness of interdisciplinary and team-based training in alternative models of healthcare delivery that may be extrapolated to more conventional models. Important research is beginning to emerge from these settings that may prove valuable to the broader context of systems-based medical education. Gilkey and Earp¹³ published original findings from their 2004 cross-sectional study, which examined qualitative perceptions and quantitative measures of integration and collaboration between volunteers representing various professional disciplines staffing a student-run health clinic. From these findings, they drew strategic conclusions that they identified as 5 essential elements for training students in interdisciplinary training programs: (1) increasing the frequency of volunteer participation, (2) enhancing interdisciplinary training and encouraging dialogue within the context of the clinic with respect to the various roles of each professional discipline, (3) increasing face time between participants, (4) building interdisciplinary relationships through various structured and unstructured settings including but not exclusive to team meetings, and (5) sharing perceptions about inclusiveness and the roles of each discipline to encourage critical thinking about how to improve relations in the context of clinical care. Their ultimate conclusion was that despite a formidable backdrop of interdisciplinary care, SRCs need to take an active role in promoting such relationships as the very professional stereotypes that they mean to dispel may actually be promoted in such well-intentioned endeavors.

Resource Acquisition and Allocation

Simpson and Long's pioneering study¹² identifies mean annual operating budgets for the respondents

of their survey of approximately \$19,000. Considerable variability exists, depending on what services are offered, which members work voluntarily and which are paid, and how much institutional support is garnered both monetarily and in personnel, supplies, laboratories, referrals, and space.^{23,26} Just under 80% of the clinics surveyed by Simpson and Long took no payments for services; additionally, many of these clinics offered pharmaceuticals on site. How and whether students are navigating cost subsidies through grant and philanthropic funds, negotiation with pharmacies, and charitable relationships with pharmaceutical companies can in large part only be surmised but not definitively determined. In 2000–2007, the American Association of Medical Colleges awarded at least 36 Caring for Community Grants to individual SRCs; these grants by definition must be authored primarily by students to fund a student-directed service endeavor.⁴⁴ This aspect of money acquisition, negotiation, and distribution connotes a large degree of money handling and control in the hands of students that directly affects the quality of care for patients served by SRCs. Limitations of funds and resources mean a shift in the reliance of students on tests to other more conventional and more accessible resources, that is, the physical examination and curbside dialogues with specialists, to inform clinical decision making. In 1995, the Asian Clinic at UC Davis revealed that simple diagnostic services were markedly limited; instead, “carefully taken histories, physical examinations, and clinical acumen lead to diagnosis and treatment of patient ailments.”¹⁸ One of the central accomplishments touted by the Camillus Health Concern, the SRC of the University of Miami School of Medicine, is promotion of an “environment that fosters clinical judgment rather than expensive tests.”⁴³

Many of these clinics offer referrals to laboratories and specialty services; the UCSD project, for instance, purchases laboratories at such reduced cost that it can offer them to clientele at no out-of-pocket expense.²³ However, the clinic has mandated that patients receive laboratory draws only on their second visit to ensure commitment to continuous care at the clinic. This is a potent example of resource allocation of services to only committed patients and of a clinic’s assurance of quality care for patients by insisting on a displayed commitment by each patient to ongoing self-care.

Substantial variety exists across clinics as to whether they offer specialty care, what kind of specialty care is offered, and whether this occurs on or off site.^{12,23} Ultimately, in the case of the UCSD experience, specialty referral acquisition was determined by the students in charge of the clinic.²³

Thus, selection and allocation of resources require students running SRCs to exert a great deal of insight, judgment, and system manipulation to acquire the right kinds of services that will adequately meet their patients’ needs and fit the constructs of the clinic.

Quality assessment of resources requires knowledge of the health status of patients served by a clinic. Cadzow and colleagues²⁹ offered the most compelling example of the efforts of an SRC to understand the population it serves. Through a health-risk questionnaire, students obtained comprehensive demographic, chronic disease, and mental health data on their patient population. They noted the relevance of a study like this to better service their own community; fine-tuning resources, for instance, would allow targeting of services to determined at-risk patients. Additionally, adapting the education of volunteer students to familiarize them with highly prevalent disorders and socioeconomic stressors particular to their population would enhance care at their SRC. Recognizing the heterogeneity of communities served by SRCs, the authors advocate for a larger application of their study to help other SRCs to enhance their need assessment efforts through population studies and information sharing.

Advocacy and Navigation of the Healthcare System to Overcome Barriers

Navigating the healthcare system requires an understanding of insurance benefits, qualifications, and system loopholes and barriers that prevent uninsured patients from obtaining necessary resources. The SRC presents a distinctive opportunity for students to advocate for patient benefits through direct navigation of the insurance enrollment process. Niescierenko and colleagues²⁸ describe the Light-house Insurance Initiative, a student-facilitated benefits enrollment process in which student volunteers from their free clinic in Buffalo, NY screen and facilitate qualified uninsured patients and their families through the insurance application process. Although other SRCs also offer benefits enrollment as part of their services,^{23,24} this study represents the first in the literature to present eligibility data and the success of the clinic in enrolling patients; in a single year, nearly 72% of their uninsured patients were screened, and nearly 26% of their uninsured population was found to be eligible. This study underscores the more expansive role of medical students in an SRC that stretches beyond mere clinical data gathering as happens in more traditional medical training settings;

more importantly, it highlights a hands-on approach to understanding the barriers to successful benefits enrollment that not only informs students about the process but also positions students as advocates for more wide-ranging public outreach projects that may improve the insurance numbers of potentially eligible families.²⁸

Navigating the healthcare system in providing a comprehensive community resource, however, also means that students involved in SRCs should aim to understand existing resources and create meaningful community partnerships so that services are not duplicated but rather fulfill a true need. Although multiple clinics stress community partnerships as integral to their organizations,^{23,27,32} it is not clear from the existing literature how community partnerships are being used to the maximum benefit of SRC patients. Certainly, understanding voids in community resources is critical to the foundation of several of these clinics^{13,16,23,27}; only the University of Washington's Students in the Community clinic, however, specifically notes that connection of patients with community resources is key to the endeavor.²⁷ To cement the utility of SRCs in community networks, future literature will need to be more robust and explicit about whether SRCs are establishing community partnerships and how SRCs are supplementing existing voids in healthcare.

Quality Care

Comprehensive literature on the quality of care at SRCs is still lacking; to date, we know of only a single abstract that documents the efforts of a single clinic in showing that an SRC may be successful at providing uninsured persons with diabetes care that is equal or superior to that of persons with diabetes receiving public or private insurance.³¹ Using indicators commonly used to evaluate Medicaid and commercial health plans as well as accepted practice guidelines, Ryskina and colleagues³¹ found that diabetes care at Mount Sinai's student-run health clinic in East Harlem, NY was comparable to or better than state and national averages across several parameters. This small study highlights the important contribution that SRCs can make to improving the quality of care for uninsured populations and also demonstrates that students at SRCs are being exposed to the highest standards of care despite restricted resources. Arguably, for SRCs to obtain credibility and acknowledgment of fairness in the landscape of healthcare as well as medical education, more extensive research on the quality of care at SRCs will need to emerge.^{26,42}

CONCLUSIONS AND RECOMMENDATIONS FOR RESEARCH

From the dramatic rise in SRCs over the past 2 decades, one may conclude that medical students have taken control of their own education and the social responsibility of the academic medical institution to share the burden in caring for the uninsured with other safety-net providers.^{45,46} Although academic medical centers historically have been entrusted with the delivery of care to the most underserved in this country, absorbing nearly one-third of the nation's uncompensated care,⁴⁷ severe Medicare cuts to graduate medical education and the dramatic penetration and overhaul of teaching hospitals by managed care have crippled the ability of academic medical centers to persist in this service role.⁴⁸ The stress of caring for the underinsured at academic institutions is critically inopportune as the number of uninsured persons grows by over 1 million per year.⁴⁹ The adverse downstream effects of being underinsured and uninsured on a person's health has been well documented in the literature.^{50–53} These studies have formed the basis of the Institute of Medicine's recent conclusion that access as defined by insurance status or the ability to pay is one of the most important predictors of the quality of healthcare across races and ethnicities.³ Leading health policy experts have advised medical schools to push current curricula beyond merely exposing the inequities in healthcare by implementing and testing interventions that may shrink disparities in healthcare delivery at their home institutions.⁵⁴ The SRC may be seen as one type of intervention with the very goal of reinstating equitable healthcare within academic medical institutions.

Improving the quality of healthcare across various delivery systems is one of the central missions of the ACGME in instituting competencies for graduate medical education. Although achieving these competencies is not explicitly central to the mission of SRCs, it may be ultimately what, among many other objectives, these endeavors are achieving at the undergraduate level of training through a distinct yet critical safety-net delivery system.⁷

Systems-based practice is one competency that may be successfully accomplished by students operating in more autonomous, experiential learning arenas such as the SRC, despite or, more compellingly, because of its unconventional learning context; however, more research is needed to understand if the competency is being achieved more explicitly through students skills and behaviors in this and other healthcare settings.

It is our hope that future research addressing the value of SRCs in the educational arena will only increase institutional buy-in, support, and acknowledgment for SRCs as unequivocally important safety-net and educational student venues. To achieve that goal, we suggest the following issues in need of future research:

- Clearer designation of SRCs as service-learning venues with defined learner outcomes and competencies. The literature on service learning in medical education is just emerging.^{55–57} Although SRCs may harbor components that support these venues as true service-learning endeavors, these have yet to be elucidated in the literature. Valid components include self-reflection activities and structured instruction and evaluation with clear goals and objectives that support the SRC as fostering community-clinic partnerships and developing students as advocates for vulnerable populations.^{56,57} Broad outcomes that could be examined are changes in medical student knowledge and understanding of the landscape of healthcare and the implications for the uninsured; changes in the attitudes toward and beliefs about public service and access to healthcare; the evolution of understanding of the need for community partnerships in the provision of healthcare; and the implications of medical student involvement in an SRC to future care for the indigent.
- As an adjunct to this, increased qualitative and quantitative data on the mastery and wider applicability of skills and competencies gained through an SRC experience. Studies should ask whether the experience at an SRC offers significantly new perspectives and abilities in patient care and, more specifically, in minimizing healthcare disparities, augmenting navigation of systems, humanizing the treatment of disadvantaged persons, and promoting interdisciplinary relationships beyond what is offered in traditional medical arenas. Research should also address the potential wider applications of these skills in other present and future clinical arenas.
- More extensive research on the quality of care at SRCs for highly prevalent chronic diseases among the uninsured, such as cardiovascular disease, diabetes, and cancer. Specifically, research should attempt to answer how SRCs compare with clinics servicing privately and/or publicly insured populations.
- More detailed descriptive data on the demographics and clinical characteristics of student-run free clinic populations with the specific goals of advocating for resources and tailoring educational goals for

students. Are students, for instance, experiencing increased exposure to chronically ill patients at SRCs in comparison with more conventional arenas?

- Increased descriptive data on the roles of students, levels of autonomy, and degree and quality of faculty guidance at SRCs as well as sharing of data on individual programs being developed and implemented at SRCs.

ACKNOWLEDGMENTS

The authors thank past and present student and physician volunteers, leaders, and cofounders of the East Harlem Health Outreach Partnership.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Agrawal JR, Huebner J, Hedgecock J, et al. Medical students' knowledge of the U.S. health care system and their preferences for curricular change: a national survey. *Acad Med* 2005; 80: 484–488.
2. Institute of Medicine Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy of Sciences Press; 2001.
3. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academy of Sciences Press; 2002.
4. Institute of Medicine. *Improving Medical Education: Enhancing the Behavioral and Social Context of Medical School Curricula*. Washington, DC: National Academy of Sciences Press; 2004.
5. Rabinowitz HK, Babbott D, Bastacky S, et al. Innovative approaches to educating medical students for practice in a changing healthcare environment: the national UME-21 project. *Acad Med* 2001; 76: 587–597.
6. Rivo ML, Keller DR, Teherani A, et al. Practicing effectively in today's health system: teaching systems-based care. *Fam Med* 2004; 36(suppl): S63–S67.
7. Accreditation Council on Graduate Medical Education. The outcome project and its six general competencies. Available at: <http://www.acgme.org/outcome>. Accessed July 2008.
8. Yap OW, Thornton DJ. The Arbor Free Clinic at Stanford: a multidisciplinary effort. *JAMA* 1995; 273: 431.
9. O'Connell MT, Rivo ML, Mechaber A, Weiss BA. A curriculum in systems-based care; experiential learning changes in student knowledge and attitudes. *Fam Med* 2004; 36(suppl): S98–S104.

10. Geller S, Taylor BM, Scott HD. Free clinics helping to patch the safety net. *J Health Care Poor Underserved* 2004; 15: 42–51.
11. Poulsen EJ. Student-run clinics: a double opportunity. *JAMA* 1995; 273: 430.
12. Simpson SA, Long JA. Medical student-run health clinics: important contributors to patient care and medical education. *J Gen Intern Med* 2007; 22: 352–356.
13. Gilkey MB, Earp JAL. Effective interdisciplinary training: lessons from the University of North Carolina's Student Health Action Coalition. *Acad Med* 2006; 81: 749–758.
14. Nordling MK. Starting a student-run homeless clinic. Available at: <http://www.amsa.org/pdf/homelessclinic.pdf>.
15. Campos-Outcalt DE. Specialties chosen by medical students who participated in a student-run, community-based free clinic. *Am J Prev Med* 1985; 1: 50–51.
16. Cohen J. Eight steps for starting a student-run clinic. *JAMA* 1995; 273: 434–435.
17. Collins AC. The Hahnemann Homeless Clinics Project: taking health care to the streets and shelters. *JAMA* 1995; 273: 433.
18. Pi R. The Asian Clinic at UC Davis: serving a minority population for two decades. *JAMA* 1995; 273: 432.
19. Haq CL, Cleeland L, Gjerde CL, et al. Student and faculty collaboration in a clinic for the medically underserved. *Fam Med* 1996; 28: 570–574.
20. Davenport BA. Witnessing and the medical gaze: how medical students learn to see at a free clinic for the homeless. *Med Anthropol Q* 2000; 14: 310–327.
21. Steinbach A, Swartzberg J, Carbone V. The Berkeley Suitcase Clinic: homeless services by undergraduate and medical student teams. *Acad Med* 2001; 76: 524.
22. Clark DL, Melillo A, Wallace D, et al. A multidisciplinary, learner-centered, student-run clinic for the homeless. *Fam Med* 2003; 35: 394–397.
23. Beck E. The UCSD student-run free clinic project: transdisciplinary health professional education. *J Health Care Poor Underserved* 2005; 16: 207–219.
24. Bennard B, Wilson JL, Ferguson KP, Sliger C. A student-run outreach clinic for rural communities in Appalachia. *Acad Med* 2004; 79: 666–671.
25. Hurtado-Rogers S, Smith E, Meah YS. Education through service: a medical student-facilitated benefits enrollment initiative. Paper presented at: SGIM National Conference; 2005; Los Angeles, CA.
26. Buchanan D, Witlen R. Balancing service and education: ethical management of student-run clinics. *J Health Care Poor Underserved* 2006; 17: 477–485.
27. Moskowitz D, Glasco J, Johnson B, Wang G. Students in the community: an interprofessional student-run free clinic. *J Interprof Care* 2006; 20: 254–259.
28. Niescierenko ML, Cadzow RB, Fox CH. Insuring the uninsured: a student-run initiative to improve access to care in an urban community. *J Natl Med Assoc* 2006; 98: 906–911.
29. Cadzow RB, Servoss TJ, Fox CH. The health status of patients of a student-run free medical clinic in inner-city Buffalo, NY. *J Am Board Fam Med* 2007; 20: 572–580.
30. Hastings J, Zulman D, Wali S. UCLA mobile clinic project. *J Health Care Poor Underserved* 2007; 18: 744–748.
31. Ryskina KL, Meah YS, Wong M, et al. Quality of diabetes care at a student-run community health clinic: how does free clinic care compare to publicly and privately insured populations? Paper presented at: SGIM Annual Meeting; April 2008; Pittsburgh, PA.
32. Jimenez M, Tan-Billet J, Babineau J, et al. The promise clinic. A service learning approach to increasing access to health care. *J Health Care Poor Underserved* 2008; 19: 935–943.
33. Kaufman DM. ABC of learning and teaching in medicine: applying educational theory in practice. *BMJ* 2003; 326: 213–216.
34. Whitman N. A review of constructivism: understanding and using a relatively new theory. *Fam Med* 1993; 25: 517–521.
35. Spencer JA. Learner-centred approaches in medical education. *BMJ* 1999; 318: 1280–1283.
36. Knowles M. *The Adult Learner: A Neglected Species*. Houston, TX: Gulf Publishing; 1990.
37. Kolb DA. *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall; 1984.
38. Hudson JN. Integrating the art and science of medicine at the early stages of medical training. *Med Educ* 2000; 34: 954.
39. Williams GC, Deci EL. The importance of supporting autonomy in medical education. *Ann Intern Med* 1998; 129: 303–308.
40. Academic Society for Student-Run Healthcare Clinics. Available at: <http://www.studentrunclinic.org>. Accessed October 2008.
41. Eckenfels EJ. Contemporary medical students' quest for self-fulfillment through community service. *Acad Med* 1997; 72: 1043–1050.
42. Dodson JA, Keller AS. Medical student care of indigent populations. *JAMA* 2004; 291: 121.
43. Fournier AM, Perez-Stable A, Greer PJ Jr. Lessons from a clinic for the homeless. The Camillus Health Concern. *JAMA* 1993; 270: 2721–2724.
44. AAMC Caring for Community. Available at: <http://www.aamc.org/about/awards/cfc07.htm>. Accessed August 2008.
45. Nadkarni MM, Philbrick JT. Free clinics and the uninsured: the increasing demands of chronic illness. *J Health Care Poor Underserved* 2003; 14: 165–174.
46. Nadkarni MM, John T. Free clinics: a national survey. *Am J Med Sci* 2005; 330: 25–31.
47. Schroeder SA. Understanding health behavior and speaking out on the uninsured: two leadership opportunities. *Acad Med* 1999; 74: 1163–1171.
48. Cunningham PJ, Grossman JM, St Peter RF, Lesser CS. Managed care and physician's provision of charity care. *JAMA* 1999; 281: 1087–1092.
49. US Census Bureau. Health insurance. Available at: <http://www.census.gov/hhes>. Accessed August 2008.
50. Ayanian JZ, Kohler BA, Abe T, Epstein AM. The relation between health insurance coverage and clinical outcomes among women with breast cancer. *N Engl J Med* 1993; 329: 326–331.
51. Ayanian JZ, Weissman JS, Schneider EC, et al. Unmet health need of uninsured adults in the United States. *JAMA* 2000; 284: 2061–2069.
52. Canto JG, Rogers WJ, French WJ, et al. Payer status and the utilization of hospital resources in acute myocardial infarction: a report from the National Registry of Myocardial Infarction 2. *Arch Intern Med* 2000; 160: 817–823.

53. Kelz RR, Gimotty PA, Polsky D, et al. Morbidity and mortality of colorectal carcinoma surgery differs by insurance status. *Cancer* 2004; 101: 2187–2194.
54. Betancourt JR, Maina AW. The Institute of Medicine report “Unequal Treatment”: implications for academic health centers. *Mt Sinai J Med* 2004; 71: 314–21.
55. Cashman SB, Seifer SD. Service-learning: an integral part of undergraduate public health. *Am J Prev Med* 2008; 35: 273–278.
56. Borges NJ, Hartung PJ. Service learning in medical education: project description and evaluation. *Int J Teach Learn Higher Educ* 2007; 19: 1–7.
57. Burrows MS, Chauvin A, Lazarus CJ, Chehardy P. Required service learning in medical students: program description and student response. *Teach Learn Med* 1999; 11: 223–231.

Medical Student Documentation in the Medical Record: A Liability?

Peter Gliatto, MD,¹ Philip Masters, MD,² and Reena Karani, MD¹

¹Mount Sinai School of Medicine, New York, NY

²Penn State College of Medicine, Hershey, PA

ABSTRACT

Medical students have routinely documented patient encounters in both inpatient and outpatient care venues. This hands-on experience has provided a way for students to reflect on patient encounters, learn proper documentation skills, and attain a sense of being actively involved in and responsible for the care of patients. Over the last several years, the practice of student note writing has come into question. Institutional disincentives to student documentation include insurance regulations that restrict student documentation from substantiating billing claims, concerns about the legal status of student notes, and implementation of electronic medical records that do not allow or restrict student access. The increased scrutiny of the medical record from pay-for-performance programs and other quality measures will likely add to the pressure to exclude students from writing notes. This trend in limiting medical student documentation may have wide-ranging consequences for student education, from delaying the learning of proper documentation skills to limiting training opportunities. This article reviews the educational value of student note writing, the factors that have made student documentation problematic, and the potential educational impact of limiting student documentation. In addition, it offers some suggestions for future research to guide policy in this area. *Mt Sinai J Med* 76:357–364, 2009. © 2009 Mount Sinai School of Medicine

Key Words: chart note, documentation, electronic medical record, medical record, medical student, progress note.

Graduating medical students are at the front lines of caring for patients at academic medical centers and must be able to communicate effectively with patients, families, and members of the healthcare team. Chart or visit notes are forms of such communication and are required parts of the legal medical record. Over the past several years, however, the practice of having students write notes in the medical record has become problematic. Consider the following scenarios that will likely sound familiar to medical educators who work with students:

- An academic medical center institutes an electronic medical record (EMR). When discussing how to handle documentation by medical students, a committee, citing legal concerns, decides that students can have read access to the system but cannot enter any data into it.
- A community hospital that hosts students from several schools is preparing for a review by the Joint Commission. Concerned about students potentially using inappropriate abbreviations that could lead to citations and the difficulty of auditing students from different schools, the hospital leadership decides to have students place their notes in a separate section of the chart that will be removed and destroyed when a patient is discharged. The hospital continues this policy after the Joint Commission audit is over.
- A school is expanding its class size and is recruiting additional community physicians to teach students in their offices as part of an ambulatory care clerkship. A large multispecialty practice that could potentially accommodate several students declines the offer because of concerns that students could adversely affect productivity. When the clerkship director calls the office, one of the senior partners explains that students are a “time drain” because “none of their work counts.”

Documenting clinical encounters in the medical record serves many purposes. Chart notes record clinical data, decisions, and procedures, facilitate

Address Correspondence to:

Peter Gliatto, MD

Mount Sinai School of Medicine
New York, NY

Email: peter.gliatto@mssm.edu

communication among health providers, support claims for payment by third-party payers, and are admissible as evidence in litigation.¹ Incomplete, inaccurate, or idiosyncratic (eg, using inappropriate abbreviations) documentation can have adverse effects on patient care or result in charges of fraud or other legal action. Therefore, it is essential that physicians pay meticulous attention to their documentation practices.

Medical schools and residency training programs have a duty to teach proper documentation skills. Yet concerns about improper documentation by students or the fact that physicians may not use student notes to support billing claims² has threatened to curtail students' ability to enter information into the medical record. These restrictions have called into question how students learn proper documentation skills and the role students should play in the clinical environment.³

This article reviews the literature on the educational value and accuracy of student documentation, the rationale behind restrictions on student documentation, and the potential impact of the restrictions on student education. The goals of this review are to stimulate a dialogue, identify future areas of research, and consider the policy implications of this important topic.

CHART DOCUMENTATION BY MEDICAL STUDENTS

The process of learning to document patient encounters serves an essential function in the training of physicians. Both the Association of American Medical Colleges (AAMC) and the Accreditation Council for Graduate Medical Education (ACGME) have specified that the ability to communicate through writing is an important educational outcome for both undergraduate and graduate medical trainees.^{4,5} The AAMC's Learning Objectives for Medical Student Education specify that medical students should be able to "communicate effectively, both orally and in writing, with patients, patients' families, colleagues, and others with whom physicians must exchange information in carrying out their responsibilities."⁴ Neither the AAMC nor the ACGME specifies what written skills trainees should have or by what educational level they should be able to demonstrate these skills. The United States Medical Licensing Examination, by means of the Step 2 Clinical Skills Examination, does evaluate the ability of trainees to document standardized clinical encounters.

Educational Value and Effective Teaching Methods of Student Note Writing

Although learning proper documentation plays such a central role in medical training, the literature on specific educational activities associated with high-quality learning in the area of medical student note writing is sparse. One study of 170 third-year medical students at a single academic medical center found that, among other things, receiving feedback on the written history and physical examination and on daily progress notes was associated with students' perceptions of high-quality teaching.⁶ Despite the paucity of data, having students participate in the documentation of the patient care that they witness and engage in is a sound educational activity from the perspective of learning theory.⁷ Principles of learning relevant to student note writing include that all education comes through experience (Dewey) and that the active process of doing and reflecting facilitates enduring learning (Lave and Wenger). In addition, learning is enhanced when the subject matter is relevant and the learner feels motivated to learn (Maslow). Lastly, by being offered instructional activities that promote increased independence and include opportunities for self-direction and reflection, learners are able to define their learning goals and monitor their progress in achieving them.

The multiple uses of student note writing in the educational process overlap considerably and change progressively across the 4 years of medical school. In the preclinical years, learning to record clinical data serves as a framework for teaching and assessing essential skills such as history taking. As students are introduced to the direct care of patients in the clinical years of medical school, the focus of documentation shifts to their ability to prioritize, synthesize, and integrate clinical information. The process of writing notes on their patients compels students to organize their thinking and to express and examine their own reasoning processes. Likewise, educators use student documentation to assess a learner's fund of basic science and clinical knowledge and the ability to apply these data to the management of clinical problems of increasing complexity.

Later in the training process, in the fourth year and beyond, student documentation takes on the more explicit purpose of conveying essential information from one clinician to another. Although this aspect of note writing appears obvious, changes in the structure of medical training in recent years, such as work-hour restrictions and the associated additional handoffs of patient care, have made the ability to effectively exchange information a critical skill that must be mastered to safely care for patients

as students proceed through training.⁸ Additionally, having students document their patient encounters provides them with a sense of ownership and an ability to contribute to the work of a team or an office,⁹ thereby fostering a sense of professional responsibility.

Quality and Accuracy of Medical Student Documentation

Are students accurate and effective documenters? The published evidence is scant and mixed. Usatine *et al.*⁹ observed 3 family practitioners on occasions when students were and were not present and compared the quality of students' notes versus physicians' notes. They found that students' notes were more legible and in general more comprehensible than when physicians were the sole documenters. All student notes were signed by physicians. Physicians added to and/or corrected most of the notes; 21% had 1 addition, 26% had 3 to 4 additions, and 11% had corrections. On several measures, student notes were more complete, notably in the documentation of follow-up plans and patient education. Attending physicians were superior in documenting a diagnosis or differential diagnosis, which occurred in 58% of attending physician notes versus 42% of student notes. This study was small, involving 3 physicians, 6 students, and 31 patient encounters, and was not blinded or randomized. Additionally, the complexity or acuity of the patients was not factored into the analysis. However, this study does suggest that students can write legible and comprehensible notes. Also, the fact that attending physicians were superior at documenting diagnoses and differential diagnoses suggests that the feedback that students receive about their notes can facilitate learning data synthesis and clinical reasoning.

On the most basic level, documentation of patient encounters needs to be accurate. The accuracy of the documentation of a clinical encounter is dependent not just on the skills of organizing, weighing, synthesizing, and recording data but also on how the data were acquired through the history and physical examination. Szauter *et al.*,¹ using standardized patient encounters, compared what senior medical students documented about the physical examinations and what the students actually did during the physical examinations, which were recorded and reviewed. Alarming, they found a high rate of inaccuracies. "Of the 207 encounter-note pairs reviewed, 169 (82%) were found to include information from examination maneuvers that had either not been performed or had been performed inaccurately." Inaccurate documentation tended to be case-specific,

and this suggests that students may have been influenced by the contexts of the cases. This study raises concerns because of its implications about the accuracy of student notes in situations with real patients. Medical teachers have limited time to do what this study did: verify documentation by direct observation of the data acquisition process. Currently, at best, doctors tend to substantiate a student's examination findings independently: "The student documented a systolic murmur; let me listen to make sure." Rarely do physicians themselves watch students with patients,¹⁰ although this could lead to much richer information: "The student is listening to the heart, but through the gown in a noisy room. Therefore, his reporting not hearing a murmur may not be accurate."

Inaccurate or inappropriate documentation is certainly not just a potential issue with student notes. A recent study of self-reported documentation practices by attending physicians revealed that over half of the physicians surveyed admitted to engaging in questionable documentation practices. For example, 40% admitted to recording laboratory data in patient charts on the basis of information that they did not personally obtain, and 6% stated that they wrote notes on patients that they did not see or examine.¹¹

Medical Students and Electronic Medical Records

Electronic Medical Records (EMRs) have increasingly been integrated into practice environments as the literature to support their use has grown and agencies such as the Institute of Medicine have called for widespread adoption.¹² Health information technology, including the promotion of EMRs, is a major focus of the healthcare spending provisions of the American Recovery and Reinvestment Act of 2009.¹³ Literature on EMRs suggests that they increase delivery of preventive care,¹⁴ enhance monitoring of drug therapy,¹⁵ and improve adherence to evidence-based guidelines.¹⁶

EMRs can also be used to profile clinical experiences.¹⁷ Patient demographic and clinical diagnosis data may be used to monitor trainee exposure to specific conditions and tailor subsequent efforts to address notable gaps. Furthermore, EMR databases can be queried for quality improvement measures at the level of the trainee or the practice site. Thus, EMRs may facilitate competency in practice-based learning and improvement, one of the ACGME competencies, and promote continuous quality improvement and self-reflection.

Published studies on the educational effects of EMRs have been less than robust, often relying on self-report rather than objective chart review.

A study of 33 third-year University of Kansas Medical Center medical students surveyed after they completed a required ambulatory medicine/family medicine clerkship found that 72% reported asking more history questions because of EMR prompts, 70% felt that essential information was easier to find electronically, 69% felt that the EMR improved their documentation, and 40% noted that they received more feedback on their EMR notes versus paper chart notes.¹⁸

Thus far, however, no study has examined the impact of EMRs on the clinical reasoning ability of learners. In one study of internal medicine and pediatric residents, surveyed respondents felt that EMR templates provided point of care education by teaching them about recommended preventive practices and reminding them to engage in anticipatory guidance.¹⁹ However, computerized decision support systems such as the preventive care reminders in that study (or others such as clinical alerts, drug-dosing assistance, and chronic disease management guidelines) may actually reduce learning as trainees turn to them instead of thinking through the clinical case at hand. The cut-and-paste functions in EMRs could also lead to shortcuts in reasoning and amplify irrelevant data or inaccuracies.²⁰

REASONS THAT MEDICAL STUDENT DOCUMENTATION MAY BE RESTRICTED

The increased scrutiny of the medical record for compliance with quality indicators and billing guidelines has made student note writing a potential liability for academic medical centers and clinical practices. Legal concerns, factors related to the implementation of EMRs, and the potential cost of medical student training to clinician productivity are additional reasons for restricting students' ability to document in the medical record.

Billing for Clinical Services

Although students have contributed directly to the medical record for decades, the tightening of Medicare billing requirements in the mid-1990s brought medicolegal concerns associated with trainee documentation to the forefront. The Physicians at Teaching Hospitals initiative by the Medicare Office of the Inspector General dramatically increased the level of scrutiny of the medical record, particularly student and resident notes, to support Medicare billing. Revisions of guidelines for billing for

evaluation and management services by the Health Care Financing Administration (HCFA) in both 1995 and 1997, intended to decrease fraudulent billing, substantially altered the use of student notes in clinical settings. HCFA guidelines maintain that the attending physician must personally document the majority of key aspects of a patient encounter in order to substantiate the service provided. Physicians may refer to student notes for elements of the review of systems and past, family, and social history, but student notes are not adequate as stand-alone documentation for billing purposes, regardless of the overall content or quality of the documentation.² Although these guidelines technically apply only to Medicare patients, institutions and practices have applied them to patients with Medicaid and other insurance plans.

Chapelle *et al.*,³ in reviewing Medicare's documentation guidelines, concluded that students should not document in the medical record: "The safest, least complicated method to ensure compliance with HCFA's guidelines is not to include student notes in any patient charts." They instead encourage physicians to have students document in a separate teaching file or to involve students in patient care in other ways, such as patient education. Others have reached the same conclusion when reviewing the potential impact of the guidelines.²¹

Other Institutional Measures

Additional incentives for institutions and practices to carefully scrutinize patient care documentation include pay-for-performance measures, the Joint Commission's core measures,⁸ and Medicare policies that hold institutions financially accountable for hospital-acquired complications.²² Anecdotally, we are aware of hospitals that have limited student documentation to avoid Joint Commission citations for use of inappropriate abbreviations. All these regulatory considerations, rooted in concerns about quality of care, patient safety, and encouragement of best practices, may lead to policies that limit or bar medical student documentation in the medical record.

Medicolegal Issues

Academic medical centers and individual practices have cited the potential for student documentation to result in or complicate patient-related or institutional legal actions as a reason for limiting or excluding student notes from the medical record.²³ Specific concerns include increased potential liability based on inaccuracies in student documentation, questions

related to patient quality and/or safety based on student involvement in patient care activities, and increased complexity of defending claims with student notes in the medical record. This topic is not well addressed in the medical literature, and the broad nature of legal actions that might potentially involve medical student documentation makes it difficult to draw meaningful generalizations concerning the significance of the role that student notes either might or do play from a medicolegal perspective.

When a legally actionable event does occur, the entire medical record, including student (and resident) notes, is considered open to discovery, or legal review. The legal significance of the actual content of student documentation, however, in the context of the overall medical record is less clear. For example, there may be discrepancies among students, residents, and faculty writing in the medical record in recording physical examination findings and other factual data, such as vital signs. There may also be differences in the assessment of a patient's overall clinical condition or progress or in the interpretation of the significance of specific findings. However, by nature of their training status, student documentation is not typically considered to be expert in nature with respect to more advanced clinicians, even if their assessments are factually correct and otherwise accurate. Whether these discrepancies truly have a negative effect on legal outcomes in such cases is unclear.

Restrictions Imposed by Electronic Medical Records

Instead of building systems to accommodate and qualify student notes, medical centers and practices may limit or deny student access to EMRs, thus eliminating the chance that a student note will make its way into the legal record and result in a billing violation or exposure to liability.

Another reason that EMR access for students may be restricted is the dearth of evidence examining the impact of EMRs on learning, as noted earlier. Although EMRs offer multiple tools to guide and streamline practice such as order sets, templates, computerized decision support systems, and reference links, the impact of these on learning is relatively unknown.

Concerns about Physician Time and Productivity

The increasing demands that primary care physicians face in providing preventive services,²⁴ the fact that physicians are not typically compensated for having

a student in an outpatient setting,²¹ and the time added to a clinic session²⁵ and the potential loss of clinical productivity caused by working with a student²⁶ are major disincentives to having a medical student work in an office setting. Usatine *et al.*⁹ found that attending physicians saved on average 3 minutes per patient if students were permitted to document encounters, even when they included the time required to edit the student note and give feedback. However, under the current Medicare rules, if a student does document a clinical encounter, the attending physician must redocument the relevant information, converting the practice of student note writing from a potentially time-neutral activity to one that is time-additive. If students are not able to contribute to patient care in a time-efficient manner, physicians may feel that they cannot afford to have students rotating in their practices.

The effect of students on house staff or attending time on inpatient services remains unclear. A physician may link to a note written by an intern or resident but not to the most relevant parts of a student note, so if the student did write a chart note the attending physician would need to independently document his or her findings. This likely takes more time than linking to an intern or resident note. Also, if a subintern is unable to write a note, the intern or senior resident would need to do so, and this could be a disincentive to assign subinterns patients. Yet, thus far, there are few studies to our knowledge that have investigated the effects of such restrictions on inpatient teaching environments.

EDUCATIONAL IMPACT OF RESTRICTING STUDENT DOCUMENTATION

The impact of altering the context of student note writing with respect to the medical record is unclear. Traditionally, in the clinical years, the ability to write and communicate with others about patients has served as a starting point from which students develop increasing independence and responsibility for thinking about and directly caring for patients. The potential loss of this qualitative contribution of student documentation in the context of caring for real patients (in contrast to generating practice notes or parallel documentation) raises concern among medical educators. Over 90% of respondents in a recent national survey of academic medical center deans believed that student notes should be included in the patient charts (Erica Friedman, personal communication, August 27, 2008). In a national

survey of clerkship directors, 64% were concerned that documentation guidelines would adversely affect the education of students.²⁷ Specific issues identified included loss of student independence and ability to actively participate in patient care, the time constraints imposed by the guidelines, loss of faculty, decreased morale, and decreased quality of care. Fifty-six percent of respondents reported having noticed an impact of the guidelines, including decreased teaching time, increased time needed to document, and loss of community-based faculty participation in educational programs. Usatine *et al.*⁹ anecdotally noted that "my informal discussions with medical students suggest that the Medicare guidelines lead students to feel undervalued and discourage them from practicing documentation skills."

Fihn *et al.*²⁸ examined the impact of the HCFA guidelines on the teaching in inpatient services. Attending physicians reported an overall increase in the amount of time devoted to patient care responsibilities and a decrease in the amount of time teaching. However, an analysis of house staff and student evaluations of teaching physicians before and after the implementation of the guidelines found no difference in their ratings of teaching faculty.

In an educational environment in which there are increasing class sizes, new medical schools,²⁹ and more students from offshore medical schools doing clinical rotations in US community hospitals,³⁰ medical schools will face increasing difficulty in finding and maintaining adequate training venues, especially for outpatient education training. If community physicians feel that students are threats to productivity, they will be less apt to host them in their practices.

CONCLUSIONS AND RECOMMENDATIONS

Faced with realistic concerns about fraud and liability, institutions and practices are restricting student note writing practices, but at the risk of abdicating one of the key missions of academic medical centers and of the profession—to educate the physicians of tomorrow. The issue of student note writing is an example of how core missions of academic medical centers—to provide high-quality patient care and to educate future physicians—can at times conflict with each other.³¹ Chiong,³² in his discussion of this tension inherent in academic medical centers, commented that "medical education may expose patients to risks that are not offset by the prospect of benefits to those individual patients but instead

by the prospect of benefits to other individuals." Using the example of procedures, Chiong wrote that to not allow a trainee to do a procedure whenever a more experienced physician was present "would not eliminate the risks associated with procedures performed by physicians who are still learning but would only add to these risks by preventing physicians from acquiring the skills necessary for competent medical practice in a responsible way." Likewise, if students are barred from documenting in the medical record, the basic skills of written communication may be deferred to internship or beyond, where supervision and feedback may be more focused on other skills.

Can written skills be taught primarily through simulation exercises or observed structured clinical encounters using standardized patients? Formal elements of documentation, such as the structure of notes and the use of precise descriptors and standard abbreviations, are probably best initially taught away from the medical record. Using standardized patient encounters to make students aware of the potential for misrepresentation of clinical findings¹ would likely be a powerful lesson on the importance of honest, accurate documentation. However, observed structured clinical encounters and exercises using standardized patients are time-intensive for faculty and costly. Additionally, it would be difficult for standardized teaching exercises to capture the complexity of patients with multisystem illnesses. Finally, note writing likely provides students with a feeling that they are contributing to the work of a clinical team.⁹ Therefore, simulation or observed encounters with standardized patients can complement the learning from clinical environments, but they cannot substitute.

Should student documentation stand alone when carefully reviewed, corrected, and signed by a supervising physician? This is a difficult position for educators to take given the present level of chart scrutiny, the stakes involved, and the lack of a guiding literature. On the other hand, if supervising physicians need to duplicate documentation, they will view student note writing as burdensome and expendable.

For the reasons outlined above, educators will likely find themselves increasingly in a position of defending the practice of student note writing in the medical record under any circumstances. Educators need to ensure that students can record data accurately and in compliance with guidelines. This requires that students receive formal instruction about documentation requirements and that student notes be audited for issues such as supervising physician cosignatures and the use of standard abbreviations.

Faculty need to be aware of the limits of student documentation for substantiating billing claims. EMRs should be used to facilitate proper documentation practices. Medical school leaders should work within their institutions and affiliated hospitals and practices to ensure that students have access to and can use EMRs.

Currently, policies and research to guide educators and administrators about the best practices for teaching written communication skills are lacking. On a national level, organizations such as the AAMC and the Alliance for Clinical Education can assist medical schools and academic medical centers by clearly articulating competencies for written communication skills by trainee level and by encouraging research that clarifies the educational value of student documentation. Several major questions deserve attention. Although there has been some exploration of faculty attitudes, what are trainees' perspectives on how restrictions may affect professional development and readiness for graduate training? Are students able to learn and comply with documentation guidelines and quality of care indicators? How does an EMR affect student note writing and physician supervision? How does student documentation affect cases in which the chart is subject to legal scrutiny? Such investigations could substantiate educators' concerns about the effects of curtailing student documentation and could help shape local and national policies in this area.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Szauter KM, Ainsworth MA, Holden MD, Mercade AC. Do students do what they write and write what they do? The match between the patient encounter and patient note. *Acad Med* 2006; 81: S44.
- Medicare's Teaching Physician Documentation Instructions*. Washington, DC: Association of American Medical Colleges; 1998.
- Chappelle KG, Blanchard SH, Ramirez-Williams MF, Fields SA. Off the charts: teaching students in compliance with HCFA guidelines. *Fam Pract Manag* 2000; 7: 37-41.
- Medical School Objectives Project. Report 1: Learning Objectives for Medical Student Education*. Washington, DC: Association of American Medical Colleges; 1998.
- Accreditation Council for Graduate Medical Education. Outcome project. Available at: <http://www.acgme.org/outcome>. Accessed August 2008.
- Torre DM, Simpson D, Sebastian JL, Elnicki DM. Learning/feedback activities and high-quality teaching: perceptions of third year medical students during an inpatient rotation. *Acad Med* 2005; 80: 950-954.
- Mann KV. The role of educational theory in continuing medical education: has it helped us? *J Contin Educ Health Prof* 2004; 24(suppl 1): S22-S30.
- Joint Commission. National patient safety goals 2009. Available at: <http://www.jointcommission.org>. Accessed August 2008.
- Usatine R, Landay M, Posalski J. Practice-based skills for students: superior student chart notes challenge Medicare documentation policies. *Acad Med* 2002; 77: S42.
- Holmboe ES. Faculty and the observation of trainee's clinical skills: problems and opportunities. *Acad Med* 2004; 79: 16-22.
- Sharma R, Kostis WJ, Wilson AC, et al. Questionable hospital chart documentation practices by physicians. *J Gen Intern Med* 2008; 23: 1865-1870.
- Institute of Medicine Committee on Improving the Patient Record. *The Computer-Based Patient Record: An Essential Technology for Health Care*. Washington, DC: National Academy Press; 1997.
- Steinbrook R. Health care and the American Recovery and Reinvestment Act. *N Engl J Med* 2009; 360: 1057-1060.
- Adams WG, Mann AM, Bauchner H. Use of an electronic medical record improves the quality of urban pediatric primary care. *Pediatrics* 2003; 111: 626-632.
- Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medical safety: a systematic review. *Arch Intern Med* 2003; 163: 1409-1416.
- Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 1998; 280: 1339-1346.
- Sequist TD, Sing S, Periera AG. Use of an electronic medical record to profile the continuity clinic experiences of primary care residents. *Acad Med* 2005; 80: 390-394.
- Rouf E, Chumley HS, Dobbie AE. Electronic health records in outpatient clinics: perspectives of third year medical students. *BMC Med Educ* 2008; 8: 1-7.
- O'Connell RT, Cho C, Shah N. Take note(s): differential HER satisfaction with two implementations under one roof. *J Am Med Inform* 2004; 11: 43-49.
- Hirschtick RE. A piece of my mind. Copy-and-paste. *JAMA* 2006; 295: 2335-2336.
- Fields SA, Usatine R, Stearns JA, et al. The use and compensation of community preceptors in U.S. medical schools. *Acad Med* 1998; 73: 95-97.
- Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med* 2007; 357: 1573-1575.
- Association of American Medical Colleges. Organization of student representatives. Available at: <http://www.aamc.org/members/osr/communications/newsletter/june03.pdf>. Accessed August 2008.
- Bodenheimer T, Grumbach K. *Improving Primary Care: Strategies and Tools for a Better Practice*. New York, NY: Lange Medical Books/McGraw-Hill; 2007; 13.
- Denton GD, Durning SJ, Hemmer PA, Pangaro LN. Research basic to medical education: a time and motion study of the effect of ambulatory medical students on

- the duration of general medicine clinics. *Teach Learn Med* 2005; 17: 285–289.
26. Vinson DC, Paden C, Devera-Sales. Impact of medical student teaching on family physicians' use of time. *J Fam Pract* 1996; 42: 243–250.
27. Fields SA, Morrison E, Yoder E, et al. Clerkship directors' perceptions of the impact of HCFA documentation guidelines. *Acad Med* 2002; 77: 543–546.
28. Fihn SD, Schleyer AM, Kelly-Hedrick H, Martin DB. Effects of the revised HCFA evaluation and management guidelines on inpatient teaching. *J Gen Intern Med* 2000; 15: 451–456.
29. Hemmer PA, Ibrahim T, Durning SJ. The impact of increasing medical school class size on clinical clerkships: a national survey of internal medicine clerkship directors. *Acad Med* 2008; 83: 432–437.
30. Hartocollis A. New York hospitals create outcry in foreign deal. *New York Times* August 5, 2008.
31. Albanese MA, Mejicano G, Gruppen L. Perspective: competency-based medical education: a defense against the four horsemen of the medical education apocalypse. *Acad Med* 2008; 83: 1132–1139.
32. Chiong WC. Justifying patient risks associated with medical education. *JAMA* 2007; 298: 1046–1048.

Direct Observation in Medical Education: Review of the Literature and Evidence for Validity

H. Barrett Fromme, MD,¹ Reena Karani, MD,² and Steven M. Downing, PhD³

¹Department of Pediatrics, University of Chicago, Chicago, IL

²Brookdale Department of Geriatrics and Adult Development, Mount Sinai School of Medicine, New York, NY

³Department of Medical Education, College of Medicine, University of Illinois, Chicago, IL

ABSTRACT

In 2000, the Accreditation Council for Medical Education introduced a new initiative that substantively changed the method by which residency programs are evaluated. In this new competency-based approach to residency education, assessment of performance became a main area of interest, and direct observation was offered as a tool to assess knowledge and skills. Despite being an inherent part of medical education as faculty and learners work together in clinical experiences, direct observation has traditionally been an informal and underused assessment method across all specialties. Residents and students report rarely being observed during their educational process, even though they value the experience. Reasons for this include a lack of faculty time, a lack of faculty skills, a potential stressful effect on the learner, and a perceived lack of validation of the assessment. This article examines the literature regarding the use of direct observation in medical education with a focus on validity evidence. We performed a PubMed search of articles pertaining to direct observation, using key words such as *direct observation*, *performance observation*, *clinical observation*, *students*, and *residents*. A subsequent search was conducted in known articles, focusing on variations of the term *observation* in the titles of articles and introducing the concept of clinical competence. In conclusion, direct observation is a unique and useful tool in the

assessment of medical students and residents. Assessing learners in natural settings offers the opportunity to see beyond what they know and into what they actually do, which is fundamentally essential to training qualified physicians. Although the literature identifies several threats to its validity as an assessment, it also demonstrates methods to minimize those threats. Based on the current recommendations and need for performance assessment in education and with attention paid to the development and design, direct observation can and should be included in medical education curricula. *Mt Sinai J Med* 76:365–371, 2009. © 2009 Mount Sinai School of Medicine

Key Words: clinical competence, direct observation, internship and residency, medical education, medical students, performance assessment, performance evaluation, performance observation, review, validation, validity.

In 2000, the Accreditation Council for Medical Education introduced an initiative that substantively changed the way in which residency programs are evaluated. The new initiative shifted the focus of accreditation from the educational opportunities that a residency program provides to an outcome-based measurement of what the residency program and its residents have actually accomplished. In this new approach to residency education, assessment of performance in 6 core competency domains became a major area of interest.¹ Historically, residencies and medical schools have relied on written assessment methods to determine knowledge achievement. Medical knowledge, however, is only 1 of the 6 core competencies. Patient care, interpersonal and communication skills, and professionalism are equally valued in medicine, a field that has clinical service to patients at its core. Standard assessment

Address Correspondence to:

H. Barrett Fromme, MD

Department of Pediatrics
University of Chicago
Chicago, IL

Email: hfromme@peds.bsd.uchicago.edu

methods, however, fail to adequately determine ability in these areas. To this end, both the Accreditation Council for Medical Education and the Liaison Committee on Medical Education have emphasized the importance of improved methods of assessment, including direct observation and global ratings of clinical performance.^{1,2}

Direct observation is embedded in the medical education and apprenticeship process. Students and residents are continually observed in the clinical setting by peers, faculty, and other members of the healthcare team. This frequency and familiarity allow direct observation to be an easily accessed teaching and assessment tool. Multiple studies have demonstrated the value of directly observed clinical experiences (when appropriately undertaken) in increasing learner comfort and ability in areas of physical examination and history taking.³ Direct observation also allows assessment at the highest level of Miller's pyramid. Although standard tests can assess the "knows" and "knows how" levels and standardized patients can demonstrate the "shows how" level, only in natural settings can it be seen if the resident "does" the skill.⁴ Unfortunately, this direct observation is most often done informally, without structure or time and without distinct goals. Despite evidence for the value of the authenticity of the experience, the formative nature of the interaction and feedback, and the call for its use by various specialties including, but not limited to, psychiatry,⁴ ophthalmology,⁵ family practice,⁶ emergency medicine,⁷ and internal medicine,⁸ direct observation is rarely used by faculty to formally assess residents and students. Although a recent survey of internal medicine residency programs found that direct observation by mini-clinical evaluation exercise (mini-CEX) was used in 90% of programs to evaluate patient care,⁹ it is unclear how often it is used, and there is little literature to support this trend in other specialties. According to a national survey of medical school pediatric clerkships, for example, students were evaluated by direct observation 57% of the time; in other words, it is used far less than other methods.¹⁰

Why does this discrepancy between theory and practice exist? If the Accreditation Council for Medical Education and Liaison Committee on Medical Education recommend direct observation for assessment and evaluation and thus many programs endorse the idea and, as has been demonstrated, it offers a way to evaluate core competencies that are difficult to assess through traditional means, why are so few residency programs and schools including formal direct observation in their evaluation toolbox? The answer is multifactorial. Opponents of direct observation have concerns, both logistical

and internal, about the process. Logistically, direct observation requires a significant investment from faculty who may not be skilled in observational feedback, and direct observation also requires a financial and resource commitment from programs, which is difficult given the financial challenges that institutions face today.¹¹ Of more concern to many are the internal validity threats to direct observation as an assessment method. Validity is the most important issue when it comes to assessment methods and their meanings. The concept of validity has evolved considerably over time. The contemporary view of validity is that all validity is construct validity, and it generally relies on 5 major sources of evidence: the scientific evidence linking the content domain to the content tested, the response process, the internal structure, the relationship to other variables, and the consequences.¹² For every assessment, there must be evidence that supports the interpretation of the results or else the results can have no true meaning. For example, if a class is given a multiple-choice examination for which there is no validity evidence, then there is no significance to passing or failing scores. For direct observation, there are many potential threats to validity due to attributes such as the reliability of the raters, the spectrum of content that it can assess, the actual effect of the observer on performance, and other similar concerns.¹³ These can be as costly as logistical issues in creating an evaluation program that includes direct observation.

Fundamentally, these threats to validity must be addressed before time and effort are spent on the logistics of implementation. In this article, we analyze the validity issues associated with assessment by direct observation of medical students and residents. Each type of evidence will be viewed through the existing literature, and recommendations will be made for enhancing the validity evidence to increase the confidence that we can have in making meaningful inferences or interpretations from direct observation data.

METHODS

A search for articles on direct observation was conducted with the PubMed database with the following search strategy: "direct observation" or "performance observation" or "clinical observation" or "clinical performance assessment" and "clinical" or "clinics" or "ambulatory" and "student" or "students" or "trainee" or "trainees" or "residents" or "resident." The terms selected for the search were developed in consultation with two of the authors, 2 with subject expertise

(H.B.F. and R.K.). The search returned an initial set of 203 articles; they were then reviewed, irrelevant articles were removed, and this left a final set of 53 articles for potential inclusion in the literature review. Irrelevant articles included articles that did not address teaching or learning, did not have medical students or residents as the observed learners, and included modalities of assessment such as objective structured clinical examinations (OSCEs), simulations, and observations of video recordings.

After we reviewed the first set of 53 articles and noted known articles that were not returned in the first search, a subsequent search was conducted, focusing on variations of the term *observation* in the titles of articles and introducing the concept of clinical competence. The new search yielded 155 new articles. After these articles were filtered for irrelevant articles with the aforementioned criteria, another 12 articles were reviewed.

The second search relied on an initial set of articles identified by the authors with subject expertise (H.B.F. and R.K.) and this ensured that both a strong search strategy and subject expertise were employed. The second search had a relatively low return of relevant articles (12 of 155) but was conducted to ensure that a thorough review was conducted.

REVIEW OF THE LITERATURE

Content Evidence

Content evidence determines if the assessment is testing what it should be testing in a way that is appropriate. In other words, are learners being tested on what is intended to be taught? Some examples of this type of evidence in traditional assessment methods include the quality of the questions, the representativeness of the question specifications, and the use of an examination blueprint, which ensures that the content tested represents the content taught. Direct observation, by nature, uses patients and real-life examples and thus depends on the random sampling of diseases and conditions available in the clinical venue. Blueprinting emphasizes the need for test content to match learning objectives. Within the clinical setting, patients do not always match the learning objectives targeted toward diseases, management choices, and so forth, and this makes blueprinting more difficult. This may be the most difficult type of validity evidence to achieve in that sense because of the uncontrollable nature of patient presentation. However, with attention to planning and organization, real nonstandardized patients can be used for assessment.

Hamdy *et al.*¹⁴ used 4 cases with real patients for sixth-year students in Bahrain. These 4 patients were actual patients with diseases that were selected from a predetermined list of common diseases, and each student worked with similar patients. Although the authors realized that 4 cases cannot represent all possible content areas, they emphasized the importance of making an effort to fulfill content blueprints in the selection of patients. Wiles and colleagues¹⁵ noted the difficulty and limitations of allowing faculty to choose patients purely for their convenience rather than the content goals. Faculty can be encouraged, when possible, to select patients that address certain content areas instead of using patients that may be convenient for other reasons.

Of course, not all patient care skills and core competencies are disease-specific. History-taking skills, physical examination skills, procedural skills, and general competencies such as professionalism and communication skills are all necessary for practice and can be assessed across disease states. Several studies, such as the one by Lane and Gottlieb,¹⁶ pointedly chose to assess skills felt to be important and consistent across encounters. The one challenge here would be to provide enough exposure to various situations and types of patient encounters (eg, the angry patient and the depressed patient) to assess the wide range of skill subsets. However, the opportunity to assess any of these skills is unique in comparison with standard assessment methods such as written tests, oral examinations, and even standardized patient encounters.

Holmboe *et al.*¹⁷ emphasized the importance of real patients and direct observation for content validity. He noted the potential difference between how a trainee performs in a testing situation and how a trainee performs with actual patients. This difference of "show how" and "does" is a critical aspect of representativeness of the encounter and blueprinting.⁴ The ultimate goal of training and assessment is to produce physicians who can perform in an actual clinical setting. It follows then that to meet the objective of assessing a learner in the setting in which he or she is to achieve clinical competence, the assessment should take place in that same setting. In this case, the method of the assessment is as much a reflection of the content as the quality of the encounter or patient. Direct observation, as a nonmediated form of assessment, meets the goal of assessing clinical competence.

Most studies that we reviewed did not comment on the process of selection for their cases, patient encounters, or skills assessed. Program and clerkship directors should be encouraged to use real patients for assessment, while making an effort to ensure that

the patients reflect a spectrum of the competencies that are emphasized in the training program. If they make this effort, direct observation becomes a valuable tool that offers a unique method of assessing residents.

Response Process Evidence

Response process refers to the cognitive processes involved in responding to the examination as both a learner and an evaluator. Students must be familiar with the type of testing, raters must be accurate in their judgments, the calculation and determination of scores must be appropriate, and the choice of any tools used for rating (eg, checklists) must be defensible.

Observer skill is one of the obvious concerns about direct observation; determining acceptable scores, reporting scores in a timely fashion, and observer comfort are all factors that faculty must face. In one study, faculty evaluators failed to detect over two-thirds of the mistakes made on a videotaped CEX.¹¹ However, several studies have demonstrated that training for faculty raters using checklists or guided evaluations improves the comfort and performance of the raters.^{11,16,17} Holmboe and colleagues¹⁷ demonstrated that of faculty trained for direct observation, 89% felt comfortable with direct observation, whereas only 11% of the untrained felt comfortable. Another aspect of response process is the way in which the ratings are handled. Raters need to be educated about the importance of completing the evaluations promptly and submitting them.¹³ Any delay in completion or submission could affect the recollection of the events and thus the validity of the ratings. The fundamental issue is the time and cost to train faculty raters. However, with experience, rater skills will improve and add to response process evidence.

The number of observations is also an issue, as an inadequate number of observations can lead to construct underrepresentation variance. As de Haes and colleagues¹⁸ discovered, scores for the observation of fifth-year students in their first clinical clerkship were more reliable as the number of assessments increased. Kogan *et al.*¹⁹ found that the reproducibility of mini-CEX scores went up with 8 encounters. Students will have exposure to an adequate number of patients and faculty, so the challenge is to find enough well-trained faculty members to observe the encounters and to determine minimum requirements for numbers.

Student familiarity with the format and process of the evaluation is slightly less of a concern because the principle behind direct observation is to see

the student in a familiar and real setting. However, at least one study suggested that the stress of student unfamiliarity with the process could affect performance.¹⁵ It is important to educate students about the form and standards used to rate their performance and to be open about the scoring process and grade determination system. In many of the studies reviewed, student preparation was not referenced and would be an area that deserves attention in future observation development.

Finally, the choice of the tool for observational assessment is a vital aspect of developing a successful direct observation program. Depending on what the assessments are to determine (pass/fail), the reliability of a checklist versus global assessments must be determined and defended. Many studies reviewed used checklists and some used global assessments, but few elucidated why the former or the latter were selected. Wass and Jolly²⁰ did compare global assessments versus checklists for the assessment of long case observation. In a traditional long case, students interview a patient, and this is followed by case presentation and discussion with faculty. Faculty do not observe the student-patient interaction. Wass and Jolly had faculty both observe the history-taking interaction and listen to the case presentation and found that for the observation, the correlation between their scores was reliable. Future use of direct observation will be more strongly supported by a discussion of the reasons for choosing checklists versus global ratings. Depending on the stakes of the assessment, one may be better than the other, or one may be easier to develop/adapt from existing tools. In either case, a sound rationale should be used.

Overall, response process evidence can be increased with observer and student training and increased observations. The literature has demonstrated that using brief observations (as short as 5 minutes) with a 2-hour faculty training session can result in high faculty confidence and increased observations.¹⁶ There are other faculty development methods, but the ideal to achieve is one that trains ample numbers of faculty with minimal time and financial costs. Effective faculty development in the context of well-chosen tools and opportunity for observation will address most of the threats to response process evidence.

Internal Structure

Internal structure evidence deals with the actual questions of a test or, in the case of direct observation, the reliability of the scores, the item difficulty, and the generalizability. This also includes issues of bias and

fairness, which are particular threats to direct observations, where there is not a standardized patient or encounter. In a standardized written examination, inter-item correlations (how well different items measure 1 construct) must be reasonably high. In a direct observation performance examination with a checklist or global rating, it is important to demonstrate moderate to high intraclass correlations (how well different raters score the same construct), correlation of rater scores, high interrater reliability (consensus of ratings given by different raters), and high generalizability coefficients (a general form of reliability used to estimate various sources of measurement error and especially useful for observational-type data). Although there are still not an overwhelming number of studies that address these specifically, the majority of the validity evidence reviewed was of this type.

The scores themselves are key to validity. The scoring instruments must produce reliable ratings or scores and adequately measure the construct of interest. Hamdy and colleagues¹⁴ found that a direct observation clinical encounter examination (DOCEE) had high intraclass and Pearson correlations between scores as well as a high generalizability coefficient. They also showed that DOCEE was not as susceptible as theorized to biases such as restriction of range (limited use of the rating scale) and rater leniency (rater tendency to give everyone high marks). Shayne and colleagues²¹ noted similar results with a standardized direct observation assessment tool for several emergency medicine competencies. Cronbach's alpha and intraclass correlations scores were above an acceptable level for reliability. Several other authors have demonstrated similarly high intraclass correlations with scores in student CEX performance,¹⁹ student overall clerkship scores,²² and student history-taking skill scores.²³ However, other studies have shown either poor score correlation (anesthesia technical skills)²⁴ or exceedingly high correlation of inflated scores.⁸ These studies express concern regarding a lack of faculty training or a lack of validation of the tool, both of which are necessary as discussed.

Although faculty training was rarely noted in the methods sections of the articles reviewed, the majority of studies that did address training were able to demonstrate high interrater reliability. For both technical skills²⁴⁻²⁶ and clinical performance in residency or clerkships,^{14,20,21,27} interrater reliability has been achieved for direct observation in some studies. Generally, as the number of raters increases, the reliability of the assessment does as well.²⁷

For any type of performance examination, the generalizability of the cases is an essential subset of internal structure evidence. Generalizability theory methods provide estimates of how well the sample

of observations represents the domain being tested (such as communication skills or physical examination skills) and help determine the number of cases required to adequately represent the domain. Despite the importance of this type of evidence, few studies even considered it in the development of their assessment, and those that did had variable answers. In Hamdy *et al.*'s¹⁴ high-stakes student examination, students at the end of medical school were required to complete encounters with 4 real patients in 4 core specialties (medicine, surgery, obstetrics, and pediatrics) while being observed. It was determined that 4 cases with 2 raters were acceptable, whereas Wiles and colleagues¹⁵ recommended (although no actual generalizability theory was used) 12 to 16 evaluations over 4 years. Further research needs to be done to determine the ideal number of cases for direct observation.

Much internal structure evidence in the current literature focuses on the reliability of the scores and the raters. Such data will continue to strengthen the evidence to support the validity of this assessment, but there is more work to be done, especially in the area of generalizability. Ultimately, as with response process evidence, it will take an initial investment of time and resources to develop the appropriate reliability and generalizability for direct observation; however, on the basis of what has been accomplished, it can be done.

Relationship to Other Variables

Direct observation is a unique type of assessment that should supplement other types of assessment. Relationships to other variables do not have to be unidirectional (ie, correlated strongly with all other variables); they can correlate well with one other type of assessment and not as well with others. For instance, it may be important that direct observation assesses a unique quality or characteristic that written examination does not. In fact, one of the goals of using direct observation should be to test learning that written examinations do not measure, such as the communication skills of a resident. At the same time, it is equally valuable to compare the assessment to one that has previously established validity or merit. In the case of direct observation, it might be interesting to correlate the direct observation of medical students with the United States Medical Licensing Examination Step II Clinical Skills Examination or with performance on an OSCE. In studies that have compared direct observation to other measures, direct observation generally increased the correlation to other measures, while being different enough that it appeared to test other constructs. Directly observed

clinical encounters were noted to sample unique areas of achievement, while correlating at some level to multiple-choice examinations and OSCEs¹⁴ in addition to patient write-ups and clerkship scores.¹⁹ In combination with other sources of data (case discussion and note review or long-case history write-up), direct observation has shown a positive correlation with clerkship clinical scores and National Board of Medical Examiners examination scores²² as well as oral²⁷ and OSCE scores.²⁰

These data indicate that direct observation provides necessary and unique information regarding learner skills while maintaining a correlation to overall performance. This is both divergent and convergent evidence, and it offers support for the role that direct observation has in medical education assessment. Programs should look to use direct observation as a method of assessment complementary to traditional methods.

Consequences

Finally, consequential validity reflects the impact that the assessment can have on the learners, whether good or bad. There is some controversy regarding the inclusion of the consequences of assessment (consequential evidence) as a specific source of validity evidence, but no matter what the perspective is, it does offer another way of viewing possible threats to the validity of an assessment. Depending on the stakes of the examination/assessment, consequential validity can be more necessary to note. Direct observation is most often used in clerkship assessment or formative feedback. Many schools use direct observation to provide students with guidance for improvement. However, many programs recognize its potential use for summative assessment such as clerkship grades. Clerkship grades are a moderate-stakes form of assessment, and direct observation contributes to the final grade but does not solely determine the grade. The harm that can come from a poor evaluation is not severe and can be counterbalanced by other achievements; this makes direct observation acceptable from the standpoint of consequences. During this review of the literature, consequential evidence was never specifically addressed by any study. It is important for any program that incorporates direct observation into its assessment system to recognize and address the rationale for using it in pass/fail determinations. Conversely, if it is to be used for summative or low- to moderate-stakes assessments, it is important to ensure that more benefit comes from the observation than harm.

DISCUSSION

The literature reviewed here has addressed direct observation in many settings, specialties, and levels of assessment, and it is fair to say that its uses are limitless. Whether formative or summative, focused or general, longitudinal or isolated, direct observation can add to the current assessment process. The key for any program is to first determine where it can be best used for learners. Does the program not have any method for observing physical examination skills, or does it need to assess procedural skills for a surgical residency? Has it found that its residents receive poor evaluations for communications skills? Is the program in need of a high-stakes summative examination for a medical school third year, or does it want recurrent formative assessments during a 3-year residency? Any and all of these can be factors to guide the introduction of direct observation. This vision is a first step toward building a successful system and will determine the level of faculty development and resource allocation that is required.

Using direct observation is not for the faint of heart or wallet. To achieve adequate validity evidence, educators need to focus on various developmental needs for the program. Educators will need to create a blueprint for the content to be tested that is based on their determination of need. They will need to use experts in the field (or train their own faculty to be experts) and develop and/or adapt checklist tools or global assessment ratings to bring adequate internal structure evidence. Learners will need to be introduced to the system to reduce any response process threats. Faculty will need to be compensated for training and for time taken to make the observations. However, many of these are startup costs that, if a relatively stable environment is assumed, will not be ongoing. The support of the administration and even external grants can go toward this initial cost that will build a foundation for future work.

Ultimately, faculty are at the heart of this process. They are often responsible for the planning and execution of direct observation and thus will need to be adequately trained and calibrated. Although faculty have long been the observers of trainees, to address response process and internal structure evidence, they will need to be prepared for their roles and responsive to feedback. Workshops have been proven effective for improving faculty comfort and skills, and additional training in assessment for those who take on leadership roles in the curriculum is advised. Attention to this early on will be endlessly valuable throughout the introduction and evaluation of any direct observation program.

CONCLUSION

Direct observation, as noted by both governing bodies and medical educators, offers a promise of a unique and necessary assessment tool for medical trainees. Although it is not without its limitations and barriers, there is ample evidence that when designed with an eye to reducing threats to the validity, direct observation can and should be incorporated into medical education. The potential benefits of assessing future physicians in natural encounters that they will face in their careers and the broad spectrum of skills and competencies that can be addressed are far too rewarding to ignore. In the end, direct observation may best be used to supplement other assessments in moderate-stakes examinations until greater validity evidence can be provided.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Accreditation Council for Graduate Medical Education. Outcomes project (2000). Available at: <http://www.acgme.org/outcome>. Accessed February 2009.
2. Liaison Committee on Medical Education. Accreditation standards (2007). Available at: <http://www.lcme.org/functionslist.htm#teaching%20and%20evaluation>. Accessed February 2009.
3. Howley LD, William WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. *Acad Med* 2004; 79: 276–280.
4. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65: s63–s67.
5. Shea SC, Green R, Barney C, et al. Designing clinical interviewing training courses for psychiatric residents: a practical primer for interviewing mentors. *Psychiatr Clin N Am* 2007; 31: 283–314.
6. Lee AG, Carter KD. Managing the new mandate in resident education: a blueprint for translating a national mandate into local compliance. *Ophthalmology* 2004; 111: 1807–1812.
7. Wendling AL. Assessing resident competency in an outpatient setting. *Fam Med* 2004; 36: 178–184.
8. Jouriles N, Burdick W, Hobgood C. Clinical assessment in emergency medicine. *Acad Emerg Med* 2002; 9: 1289–1294.
9. Chaudry SI, Holmboe E, Beasley BW. The state of evaluation in internal medicine residency. *J Gen Intern Med* 2008; 23: 1010–1015.
10. Kumar A, Gera R, Shah G, et al. Student evaluation practices in pediatric clerkships: a survey of the medical schools in the United States and Canada. *Clin Pediatr (Phila)* 2004; 43: 729–735.
11. Holmboe ES. Faculty and the observation of trainee's clinical skills: problems and opportunities. *Acad Med* 2004; 79: 16–22.
12. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ* 2003; 37: 830–837.
13. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med* 2003; 15: 270–292.
14. Hamdy H, Prasad K, Williams R, Salih FA. Reliability and validity of the direct observation clinical encounter examination (DOCEE). *Med Educ* 2003; 37: 205–212.
15. Wiles CM, Dawson K, Hughes TA, et al. Clinical skills evaluation of trainees in a neurology department. *Clin Med* 2007; 7: 265–269.
16. Lane JL, Gottlieb RP. Structured clinical observations: a method to teach clinical skills with limited time and financial resources. *Pediatrics* 2000; 105: 973–977.
17. Holmboe ES, Hawkins RE, Huot SJ. Effect of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med* 2004; 140: 874–881.
18. De Haes JCJM, Oort FG, Hulsman RL. Summative assessment of medical students' communication skills and professional attitudes through observation in practice. *Med Teach* 2005; 27: 583–589.
19. Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Acad Med* 2003; 78: s33–s35.
20. Wass V, Jolly B. Does observation add to the validity of the long case? *Med Educ* 2001; 35: 729–734.
21. Shayne P, Gallahue F, Rinnert S, et al. Reliability of a core competency checklist assessment in the emergency department: the standardized direct observation assessment tool. *Acad Emerg Med* 2006; 13: 727–732.
22. Hasnain M, Connell KJ, Downing SM, et al. Toward meaningful evaluation of clinical competence: the role of direct observation in clerkship ratings. *Acad Med* 2004; 79: S21–S24.
23. Woolliscroft JO, Calhoun JG, Beauchamp C, et al. Evaluating the medical history: observation versus write-up review. *J Med Educ* 1984; 59: 19–23.
24. Friedman Z, Katznelson R, Devito I, et al. Objective assessment of manual skills and proficiency in performing epidural anesthesia—video-assisted validation. *Reg Anesth Pain Med* 2006; 31: 304–310.
25. Beard JD, Jolly BC, Newble DI, et al. Assessing the technical skills of surgical trainees. *Br J Surg* 2005; 92: 778–782.
26. Scott DJ, Rege RV, Bergen PC, et al. Measuring operative performance after laparoscopic skills training: edited videotape versus direct observation. *J Laparoendosc Adv Surg Tech A* 2000; 10: 183–190.
27. Dunnington G, Reisner L, Witzke D, Fulginiti J. Structured single-observer methods of evaluation for the assessment of ward performance on the surgical clerkship. *Am J Surg* 1990; 159: 423–426.

Evaluating the Impact of the Humanities in Medical Education

Andrea Wershof Schwartz, MA,¹ Jeremy S. Abramson, MD, M.M.Sc.,²
Israel Wojnowich, BA,¹ Robert Accordino, MSc,¹ Edward J. Ronan, PhD,¹
and Mary R. Rifkin, PhD¹

¹Mount Sinai School of Medicine, New York, NY

²Harvard School of Medicine, Cambridge, MA

ABSTRACT

The inclusion of the humanities in medical education may offer significant potential benefits to individual future physicians and to the medical community as a whole. Debate remains, however, about the definition and precise role of the humanities in medical education, whether at the premedical, medical school, or postgraduate level. Recent trends have revealed an increasing presence of the humanities in medical training. This article reviews the literature on the impact of humanities education on the performance of medical students and residents and the challenges posed by the evaluation of the impact of humanities in medical education. Students who major in the humanities as college students perform just as well, if not better, than their peers with science backgrounds during medical school and in residency on objective measures of achievement such as National Board of Medical Examiners scores and academic grades. Although many humanities electives and courses are offered in premedical and medical school curricula, measuring and quantifying their impact has proven challenging because the courses are diverse in content and goals. Many of the published studies involve self-selected groups of students and seek to measure subjective outcomes which are difficult to measure, such as increases in empathy, professionalism, and self-care. Further research is needed to define the optimal role for humanities education in medical training; in particular, more quantitative studies are needed to examine the impact that it may have

on physician performance beyond medical school and residency. Medical educators must consider what potential benefits humanities education can contribute to medical education, how its impact can be measured, and what ultimate outcomes we hope to achieve. *Mt Sinai J Med* 76:372–380, 2009. © 2009 Mount Sinai School of Medicine

Key Words: empathy, evaluation, humanities, literature, medical education, medical student performance, premedical, professionalism, resident performance, self-care.

Medical educators have long recognized that physicians in training require more than just an understanding of scientific principles to become successful doctors.¹ As early as the 1920s, Dr. Francis Peabody of Harvard Medical School lamented that “young graduates have been taught a great deal about the mechanism of disease, but very little about the practice of medicine or, to put it more bluntly, they are too ‘scientific’ and do not know how to take care of patients.”² Inclusion of the humanities in medical education may offer significant potential benefits to individual future physicians and to the medical community as a whole and may provide an avenue to address the concerns that Dr. Peabody raised so long ago. Some of these benefits may include creating a more diverse medical profession with students who bring various backgrounds and talents the many challenges of medicine³ and perhaps creating more open-minded doctors who are able to better relate to their patients or understand their points of view. Studying humanities may enhance medical students’ ability to communicate with patients or improve their comfort level in the clinical setting,⁴ and it may increase their ability to listen to and observe diagnostic findings from patients.^{5,6} Additionally, medical school communities may be enriched by the presence of humanities; at our own institution, SinaiArts,

Address Correspondence to:

Andrea Wershof Schwartz
Mount Sinai School of Medicine
New York, NY
Email: andrea.schwartz@mssm.edu

which coordinates student music and dance performances and art exhibits, and the *Mount Sinai Mosaic*, a student literary publication,⁷ are 2 examples of ways in which the humanities can be incorporated into student life and provide opportunities for reflection about the process of becoming a physician.

Debate remains about the definition and precise role of the humanities in medical education, whether at the premedical, medical school, or postgraduate level.⁸ Consensus is lacking in the medical community about which subjects fall under the rubric of humanities education and what the goal of their inclusion in medical education may be. For the purpose of our review, we have defined humanities education broadly as non-medical subjects that may enhance the training of physicians including literature, visual arts, performing arts, ethics, philosophy, anthropology, history, and sociology.

Even as the discussion continues, the humanities are occupy an increasingly important role in medical school admissions and curricula. Calls for reform of the premedical curriculum seek to encourage students to study more than basic science and increase their appreciation of the humanities,⁹ and some colleges offer courses or majors in the medical humanities in addition to the traditional premedical preparation.¹⁰ Undergraduate humanities and social science majors make up close to 15% of successful medical school applicants.¹¹ Medical schools around the country offer electives and required courses in the humanities,¹² and the field of narrative medicine, which uses literature and writing as a vehicle for increasing physician self-reflection and empathy, is gaining increasing prominence in both the medical literature^{13,14} and the lay press.¹⁵

Driving this trend is the recognition that the study of humanities may provide a unique avenue to positively affect physicians in training. Medical educators must consider what specific contributions humanities education can make, how its impact can be measured, and what ultimate outcomes we hope to achieve. These questions are particularly challenging because they require medical educators to reflect on how to define a successful physician and how to measure that success. This article reviews the literature on the impact of humanities education on the performance of medical students and residents and the difficulties posed by the evaluation of the outcomes of a humanities education.

MEDICAL SCHOOL AND RESIDENCY PERFORMANCE OF UNDERGRADUATE HUMANITIES MAJORS

A frequently studied concern is that students who majored in the humanities prior to medical school might be inadequately equipped to perform academically at the level of their science-focused peers. At our own institution, the Mount Sinai School of Medicine (MSSM), the Humanities and Medicine (H&M) early admissions program has for 20 years accepted humanities majors to the medical school and tracked their academic progress in comparison with their peers with traditional premedical backgrounds. These students apply to medical school as college sophomores majoring in a humanities discipline and, if accepted, complete only minimal premedical requirements and do not take the Medical College Admission Test (MCAT); this provides them with maximal flexibility during their premedical years to focus on their humanities studies. To ensure adequate exposure to premedical science, these students spend a summer prior to matriculation at the medical school, studying abridged curricula in organic chemistry and physics and participating in medical ethics courses and clinical activities. In 2000, we conducted a study of the first cohort of H&M students, who matriculated at MSSM between 1991 and 1997, and compared them to 2 matched cohorts of students accepted through the standard admission process.¹⁶ We found that H&M students were more likely to have academic difficulties during the preclinical years, but by the third year, they were, as a group, indistinguishable academically from their classmates with traditional premedical backgrounds on the basis of clerkship performance. In fact, the H&M students were overrepresented among students who earned membership in the Alpha Omega Alpha National Medical Honor Society, honors in clinical clerkships, and other academic awards. These data suggest that although medical students with majors in the humanities may not succeed as well in the preclinical curriculum in comparison with peers with more extensive science backgrounds, they do excel in their clinical clerkships, where textbooks and Petri dishes give way to real patients and clinical problem solving.

Several studies have compared humanities majors to science majors in their undergraduate medical curriculum accomplishments. Zeleznik *et al.*¹⁷ and Herman and Veloski¹⁸ studied the medical school performance of students with science and nonscience bachelor's degrees. Data were collected for students entering over a 4- or 7-year period at the same medical school. No difference was found

in yearly grade point averages, in parts I and II of the United States Medical Licensing Examination (USMLE) examinations, and in ratings in 6 required clinical clerkships between the science and non-science majors. Yens and Stimmel,¹⁹ in a study of 9 classes at 1 medical school, compared a number of variables (scores on parts I and II of the USMLE examinations, course grades for all 4 years, and membership in Alpha Omega Alpha) for students with various undergraduate majors and found that nonscience majors performed as well, or sometimes better, than their science major cohort. Study after study supports this finding that there is no significant difference in medical school performance when science and nonscience majors are compared.^{20–23} Table 1 summarizes the findings of these studies, which in aggregate suggest that students who study the humanities in college perform at least as well in medical school and residency as their colleagues who studied sciences, and perhaps have an advantage in certain settings.

The single best predictor of academic achievement in medical school and residency is previous academic performance, regardless of the discipline of study. Previous success in school most closely predicts academic achievements in medical school^{24,25} and residency ranking,²⁶ although this has not been as well validated for postgraduate performance.²⁷ Students who majored in the humanities may in fact perform better as interns when rated by supervisors on criteria ranging from clinical judgment to relationships with patients and peers,²⁸ although some studies found no difference in residency performance in comparison with science undergraduate majors.^{24,29}

Beyond medical school, predictors of performance include scores on the National Board of Medical Examiners (NBME) Step 1 and 2 examinations and clinical clerkship grades; the latter have been shown to correlate closely with success in residency programs³⁰ and are reliable predictors of intern-year knowledge and professionalism ratings,³¹ particularly at the extremes of performance.³² The strength of the relationship between medical school and residency performance varies by specialty; therefore, interpreting the data across medical specialties should be performed with caution.³³ In addition, humanities background and academic performance in medical school do not have a significant influence on whether a student chooses a primary care field,^{34–36} although students who studied humanities as undergraduates may be more likely to select psychiatry as a specialty.³⁷

INCREASING PRESENCE OF HUMANITIES IN PREMEDICAL AND MEDICAL EDUCATION

Premedical education has for decades centered around the principle that applicants to medical school require preparation in the sciences and a rigorous approach to learning.³⁸ Beginning with the 1984 Association of American Medical Colleges report on the general professional education of the physician and college preparation for medicine,³⁹ calls for reform of premedical curricula have stressed the value of studying humanities in addition to traditional scientific and mathematical courses. Certain courses, such as calculus and organic chemistry, are largely irrelevant to medical practice and are used primarily to distinguish students from one another academically in the premedical setting,^{40,41} often creating a competitive learning environment antithetical to the team-based approach so central to the practice of medicine.^{42–44} In place of these often unnecessary courses, some experts suggest that premedical curricula should include studies in the humanities, which may better prepare students for the complex realities of medicine⁹ and emphasize the moral and personal development of physicians in training.⁴⁵

At the medical school level, many curricula currently offer humanities electives or required courses, including writing, literature, music, arts, and ethics,^{46–48} and this trend appears to be increasing.¹² As of 2004, 88 of 125 American medical schools required humanities courses, and 55 offered elective courses.⁴⁹ Some medical schools also offer a concentration or minor in medical humanities for medical students.^{50,51} An entire issue of *Academic Medicine* was devoted to highlighting examples of these various programs, indicating the increasing degree to which these programs are becoming commonplace within medical education.⁵² At MSSM, humanities-focused courses include electives on writing about medicine,⁵³ literature and medicine, and music and medicine, and a required component of the internal medicine/geriatrics clerkship includes guided visits to an art museum with exercises in the art of observation; this program has been favorably reviewed by students at our institution and is gaining popularity at other medical schools.⁵⁴

POSITIVE IMPACT OF HUMANITIES EDUCATION ON MEDICAL TRAINING

The increasing presence of the humanities at the premedical and medical school levels does not

Table 1. *Predictors of Success in Medical School and in Residency.*

Medical School		
Study Group	Finding	Reference
190 medical students	Despite lower uGPAs, humanities majors more likely to earn AOA membership and formal commendations.	Stratton <i>et al.</i> ⁸¹ (2003)
808 medical students	Previous academic performance is the best predictor of medical school performance.	Bastias <i>et al.</i> ²⁵ (2000)
255 medical students	Compared to matched traditional premed humanities and science majors, humanities and medicine students had more NBME Step 1 failures but earned more graduation awards and honors and were more likely to earn AOA membership.	Rifkin <i>et al.</i> ¹⁶ (2000)
406 medical students	No difference in medical school performance between humanities/social science majors and science/math majors.	Smith ²² (1998)
200 medical students	No correlation between the quantity of undergraduate science and average preclinical performance. uGPA and MCAT scores predict NBME performance.	Hall and Stocks ²¹ (1995)
782 medical students	No difference in medical school performance based on the quantity of previous science study.	Neame <i>et al.</i> ²⁴ (1992)
812 medical students	Undergraduate science majors had higher MCAT scores than humanities majors, but no statistically significant difference in performance in the first 2 years of medical school.	Ashikawa <i>et al.</i> ²⁰ (1991)
200 medical students	No difference in uGPAs or NBME performance between degree types; humanities majors consider leaving medical school more frequently, and humanities/social science majors are more likely to choose psychiatry.	Zelevnik <i>et al.</i> ¹⁷ (1983)
735 medical students	Nonscience majors do just as well or better than science majors.	Yens and Stimmel ¹⁹ (1982)
1077 medical students	No relationship between premedical training and clinical competence.	Herman and Veloski ¹⁸ (1981)
Residency		
Study Group	Finding	Reference
1069 residents	Low third-year grades predict low ratings in knowledge and professionalism; NBME Step 1 scores predict knowledge.	Greenburg <i>et al.</i> ³¹ (2007)
645 residents	Cognitive and noncognitive factors are important in predicting future success and residency ranking.	Peskun <i>et al.</i> ²⁶ (2007)
5701 psychiatry residents	Psychiatrists scored higher on measures of verbal ability but lower on NBME.	Sierles <i>et al.</i> ³⁷ (2004)
1659 medical students	NBME scores, grades, and AOA membership led to no difference in choosing generalist fields.	Stimmel and Serber ³⁴ (1999)
104 residents	Students who study the humanities in addition to science as undergraduates perform better as interns.	Rolfe <i>et al.</i> ²⁸ (1995)
581 residents	No difference in NBME Step 2 scores, but science majors scored higher on NBME Step 1. The same residency performance was found.	Elam <i>et al.</i> ²³ (1993)
277 residents	Higher NBME scores correlate with outstanding resident evaluations. NBME scores provide reliable predictions only at extremes of performance.	Yindra <i>et al.</i> ³⁰ (1988)
185 residents	uGPA and medical school achievement are not the best predictors of residency performance; NBME II is.	Gunzburger <i>et al.</i> ³² (1987)
636 medical students	Undergraduate academic background makes no difference in career plans or residency selection.	Ferrier and Woodward ³⁵ (1983)
441 residents	The strength of the relationship between medical school performance and residency performance varies by specialty.	Gonnella and Hojat ³³ (1983)
185 residents	Intern-year performance is similar, regardless of the undergraduate major.	Woodward and Mcauley ²⁹ (1983)
135 students	Similar NBME scores and residency selection were found, and there was no adverse effect of humanities education.	Dickman <i>et al.</i> ³⁶ (1980)

NOTE: The studies are presented in reverse chronological order.

Abbreviations: AOA, Alpha Omega Alpha Medical Honor Society; MCAT, Medical College Admission Test; NBME, National Board of Medical Examiners; uGPA, undergraduate grade point average.

correlate with the measuring of its impact on physicians in training. This lack of research is partially due to the diverse reasons why medical educators turn to study of the humanities.⁵⁵ Humanities courses are used in a variety of settings at the premedical and medical school levels, with no consensus in the medical community about the desired outcomes or benefits from participation in these courses, which are often more qualitative than quantitative. Humanities education may be used as a vehicle to incorporate nonscientific aspects of medical education,⁴⁵ such as professionalism, communication,⁵⁶ cultural sensitivity, and ethics.⁴⁹ Studying the humanities may also be used to combat a perceived loss of empathy that may occur over the course of medical training.⁵⁷ In some institutions, humanities education is used, together with rituals such as White Coat Ceremonies, to formalize a commitment to the values of humanism in medicine.^{58,59}

Robust interdisciplinary programs have been developed to bring faculty and students from humanities departments together with medical faculty and students.^{60,61} In some cases, more limited programs introduce literature or poetry to focus discussion of clinical cases in an attempt to highlight humanistic aspects of patient care.⁶² However, the diverse goals of and outcomes anticipated from programs at medical schools thus make it challenging to measure what students gain from such courses. The data available fall into 3 main categories: the ability of humanities education to promote empathy, professionalism, and medical student self-care.

Levels of empathy have been noted to decline over the course of medical school, with younger students and students earlier in their training scoring higher on validated measurements of empathy such as the Jefferson Scale of Physician Empathy (JSPE).^{63,64} A prospective longitudinal study measured empathy among 125 medical students at the beginning and end of the third year of medical school using the JSPE and found statistically significant declines when comparing the 2 tests.⁶⁵ According to one study, students with higher empathy scores on the JSPE tend to earn higher ratings of clinical competence on clerkships, although scores did not correlate with performance on NBME or MCAT examinations.⁶⁶

Debate exists about whether empathy levels change during residency. A study that administered the JSPE to 98 internal medicine residents found that empathy scores remained stable throughout residency,⁶⁷ whereas a separate prospective, longitudinal study that measured residents' empathy using the Interpersonal Reactivity Index found that empathy

decreased slightly during residency. This latter study found that residents' knowledge, as measured by the Internal Medicine In-Training Examination, increased during the same time period and concluded that, because knowledge and empathy are not necessarily linked, residency programs should focus on cultivating each competency separately within training curricula.⁶⁸ Even if a decline in empathy as measured by the JSPE or Interpersonal Reactivity Index does occur during medical training, it may not be clinically significant. However, the downward trends suggest that empathy may indeed decrease during this critical period in a physician's development, raising the question of which aspects of medical training contribute to this decline and, in turn, which interventions could be effective in minimizing or even reversing this trend.

Humanities education may play an important role in cultivating or maintaining empathy during medical training. A qualitative survey by DasGupta and Charon⁶⁹ found that among 16 medical students who participated in a reflective writing workshop, several reported increases in empathy on a 7-question survey administered after the course. A quantitative/qualitative study by Shapiro *et al.*⁷⁰ examined the impact of an elective course discussing literature and medicine. The 22 students who volunteered to participate in the course were found to have improved levels of empathy after the intervention, as measured by the Empathy Construct Rating Scale. Qualitative evaluations found that the students' understanding of the patient's perspective increased in complexity, and students also noted a role for the humanities in dealing with the stress of medical training. Another study found that participation in sessions on empathy and spirituality correlated with higher scores on the JSPE.⁶⁴ Empathy and humanism ratings are highly correlated⁶⁷ and are sometimes used interchangeably when physician or medical student performance is being rated.

The second dimension of medical training that may be positively affected by humanities education is professionalism. Reliable, evidence-based methods for assessing professionalism are notably lacking, particularly because professionalism is often defined by one of its subcomponents, such as humanism, ethics, or cultural competency.⁷¹ Studying the humanities may lead to improvements in physicians' cultural competence and enhance their ability to care for diverse patient populations.⁷² Previous areas of study do not predict professional behavior: a retrospective cohort study found that professional behavior among medical students on clerkships could be predicted by student performance on standardized patient examinations and

adherence to required immunizations but not from medical school admissions materials.⁷³ A comparative study of 148 first-year internal medicine residents found that professionalism could be assessed with peer and supervisor observation-based assessments, which in turn were associated with knowledge and clinical performance.⁷⁴ Thus, few data are available to support the hypothesis that humanities education affects professional behavior, unless one considers professionalism and humanism to be linked, and thus changing one affects the other. Some argue that humanism and professionalism are not necessarily intertwined and in fact may be at odds with one another. One author defines humanism as "a universal, egalitarian ideology," whereas professionalism represents "the parochial, culturally determined practices of a particular professional group", arguing that blurring the distinction between the two may actually decrease humanistic behavior by asking medical students to leave behind humanistic values in favor of professional behavior.⁷⁵

The third area in which medical trainees may benefit from humanities education is student or physician self-care and well-being. Studying the humanities may provide much needed opportunities for self-reflection about the intensive process of becoming a physician and may ease feelings of isolation or burnout.⁶¹ Educational interventions have been shown to have a positive impact on the health habits and self-care practices of medical students, which may in turn increase their ability to be resilient to the challenges of medical school.⁷⁶ For instance, a randomized controlled trial involving 64 medical students suggested that writing about emotional topics could decrease the incidence of self-reported depressive symptoms and healthcare visits in a 3-month follow-up period.⁷⁷ Additionally, a course on humanism and professionalism for medical students that used reading and discussion to provide opportunities for self-reflection found that 73% of the students reported an increase in connectedness with their classmates, suggesting a role for improving communication and creating positive change in the way in which medical students interact with their patients.⁷⁸

THE CHALLENGE OF MEASURING THE IMPACT OF HUMANITIES EDUCATION

Despite these data suggesting that humanities education may positively affect empathy, professionalism, and self-care among medical trainees, exactly how exposure to the humanities engenders these attitudes

and skills and, therefore, the most effective type of humanities education, remains to be determined. One attempt at offering a conceptual basis for using narrative in medical education discussed particular pedagogical methods, such as creating a cognitive disequilibrium and using transformative learning, as possible mechanisms for creating a positive impact on medical students through humanities education.⁷⁹

Another problem with analyzing the impact of humanities education, whether at the premedical or medical school level, is that there are many confounders, such as gender, age, years spent pursuing other careers or degrees prior to medical training, and different types of coursework within the humanities. These different factors may affect students' interests in the humanities as well as the talents and abilities that they bring to their medical education experience. For instance, female medical students have repeatedly been shown to score higher on measurements of empathy.^{63,66} Another weakness of the few studies that have attempted to analyze the impact of humanities education is that many of these courses are electives, and so students participating are, by definition, a self-selected group. In such a setting, any observed increase in empathy or improvement in communication may be due to the possibility that students who already possess greater empathy or communication skills may be drawn to enroll in humanities courses in the first place. Furthermore, most of the studies are drawn from cohorts at individual institutions, which may have their own particular biases and values in the admissions process and criteria used to determine clinical and academic awards. Certain nonquantitative measures, such as patient satisfaction, may be entirely too subjective to be helpful markers of physician success and competence. Although it is much easier to compare objective findings such as grades and NBME scores, these more subjective qualities play an equal if not more significant role in determining the kind of physician that a given medical student will ultimately become.⁸⁰

CONCLUSION

Recent trends indicate an increasing appreciation of, and role for, humanities education in medical training. There is ample evidence that college humanities majors perform ably in medical school and residency and that medical schools can recruit students from nonscience backgrounds without hesitation with respect to their potential for academic success. Further research is needed to track the

performance of these students beyond residency to determine what impact, if any, humanities education can have on a physician's subsequent career choices and success as a doctor.

At MSSM, we are currently conducting a follow-up study of the cohort in our initial comparison of H&M students to students with traditional premedical backgrounds.¹⁶ We are gathering data about the same cohort of MSSM graduates, now 3 to 13 years out of medical school, and will quantify certain measures of their success as physicians, including awards, elections to prestigious societies, and publications. We will also assess choices of specialty and career paths, including academic medicine or private practice. Our study will assess the impact of humanities studies during college on medical professionals who have completed their postgraduate training. By matching the H&M cohort with students with traditional premedical backgrounds as well as students who majored in the humanities but did not participate in the early admissions program, we hope to compare performances of similar students who differ only by undergraduate major.

Future studies should better characterize the physician behaviors and competencies that medical educators can encourage in students through humanities education. Humanities courses in medical schools should set specific, measurable goals for their curricula and determine methods to evaluate whether those goals are met. Possible markers for these goals include academic achievements, fulfilling curricular competencies, performance on validated measures of empathy, and inclusion in the Gold Humanism Honor Society. Systematically evaluating the impact of different types of humanities education, within a definition of what it means to create successful physicians, could ultimately help determine the most important and effective types of humanities education in medical school. If such studies do provide evidence that students who study the humanities perform better in certain measurable areas than their traditional pre-medical counterparts, it would further legitimize the place of humanities education within medical education and create new ways in which to prepare students for the privilege and responsibilities of physicianhood.

DISCLOSURES

Potential conflict of interest: Mary Rifkin and Edward Ronan are Director and Associate Director, respectively, of the Humanities and Medicine Program at The Mount Sinai School of Medicine. Andrea Wershof

Schwartz, Jeremy Abramson, Robert Accordino and Israel Wojnowich were participants in the Humanities and Medicine Program at The Mount Sinai School of Medicine.

REFERENCES

1. Flexner A. *Medical education in the United States and Canada: a report to the Carnegie Foundation for the Advancement of Teaching*. New York: Carnegie Foundation for the Advancement of Teaching, 1910, page 44. Available at: http://www.carnegiefoundation.org/files/elibrary/Carnegie_Flexner_Report.pdf. Accessed June 2009.
2. Peabody F. The care of the patient. *JAMA* 1927; 88: 877.
3. Kirch DG. The gateway to being a doctor: rethinking premedical education. Available at: <http://www.aamc.org/newsroom/reporter/april08/word.htm>. Accessed May 2009.
4. Jauhar S. From all walks of life—nontraditional medical students and the future of medicine. *N Engl J Med* 2008; 359: 224–227.
5. Naghshineh S, Hafner JP, Miller AR, et al. Formal art observation training improves medical students' visual diagnostic skills. *J Gen Intern Med* 2008; 23: 991–997.
6. Shapiro J, Rucker L, Beck J. Training the clinical eye and mind: using the arts to develop medical students' observational and pattern recognition skills. *Med Educ* 2006; 40: 263–268.
7. The Mount Sinai Mosaic: a student literary publication. Available at: <http://www.mssm.edu/students/groups/mosaic>. Accessed May 2009.
8. Borek D. Unchanging dilemmas in American medical education. *Acad Med* 1989; 64: 240–244.
9. Emanuel EJ. Changing premed requirements and the medical curriculum. *JAMA* 2006; 296: 1128–1131.
10. Fried C, Madar S, Donley C. The biomedical humanities program: merging humanities and science in a pre-medical curriculum at Hiram College. *Acad Med* 2003; 78: 993–996.
11. Association of American Medical Colleges. MCAT and GPAs for applicants and matriculants to U.S. medical schools by primary undergraduate major, 2007. Available at: <http://www.aamc.org/data/facts/2007/mcatgpabymaj07.htm>. Accessed May 2009.
12. Literature, arts and medicine online database. Available at: <http://litmed.med.nyu.edu>. Accessed May 2009.
13. Charon R. Narrative and medicine. *N Engl J Med* 2004; 350: 862–864.
14. Charon R. The patient-physician relationship. Narrative medicine: a model for empathy, reflection, profession, and trust. *JAMA* 2001; 286: 1897–1902.
15. Chen PW. Stories in the service of making a better doctor. Available at: <http://www.nytimes.com/2008/10/24/health/chen10-23.html>. Accessed May 2009.
16. Rifkin MR, Smith KD, Stimmel BD, et al. The Mount Sinai humanities and medicine program: an alternative pathway to medical school. *Acad Med* 2000; 75(suppl): S124–S126.
17. Zeleznik C, Hojat M, Veloski J. Baccalaureate preparation for medical school: does type of degree make a difference? *J Med Educ* 1983; 58: 26–33.

18. Herman MW, Veloski JJ. Premedical training, personal characteristics and performance in medical school. *Med Educ* 1981; 15: 363–367.
19. Yens DP, Stimmel B. Science versus nonscience undergraduate studies for medical school: a study of nine classes. *J Med Educ* 1982; 57: 429–435.
20. Ashikawa H, Hojat M, Zeleznik C, Gonnella JS. Reexamination of relationships between students' undergraduate majors, medical school performances, and career plans at Jefferson Medical College. *Acad Med* 1991; 66: 458–464.
21. Hall ML, Stocks MT. Relationship between quantity of undergraduate science preparation and preclinical performance in medical school. *Acad Med* 1995; 70: 230–235.
22. Smith SR. Effect of undergraduate college major on performance in medical school. *Acad Med* 1998; 73: 1006–1008.
23. Elam CL, Johnson MM, Johnson R. Students' premedical preparations and academic performances in medical school and residency. *Acad Med* 1993; 68: 229–230.
24. Neame RL, Powis DA, Bristow T. Should medical students be selected only from recent school-leavers who have studied science? *Med Educ* 1992; 26: 433–440.
25. Bastias G, Villarroel L, Zuniga D, et al. Academic performance of medical students: a predictable result [in Spanish?]. *Rev Med Chil* 2000; 128: 671–678.
26. Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Med Educ* 2007; 41: 57–64.
27. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002; 324: 952–957.
28. Rolfe IE, Pearson S, Powis DA, Smith AJ. Time for a review of admission to medical school? *Lancet* 1995; 346: 1329–1333.
29. Woodward CA, Mcauley RG. Can the academic background of medical graduates be detected during internship? *Can Med Assoc J* 1983; 129: 567–569.
30. Yindra KJ, Rosenfeld PS, Donnelly MB. Medical school achievements as predictors of residency performance. *J Med Educ* 1988; 63: 356–363.
31. Greenburg DL, Durning SJ, Cohen DL, et al. Identifying medical students likely to exhibit poor professionalism and knowledge during internship. *J Gen Intern Med* 2007; 22: 1711–1717.
32. Gunzburger LK, Frazier RG, Yang LM, et al. Premedical and medical school performance in predicting first-year residency performance. *J Med Educ* 1987; 62: 379–384.
33. Gonnella JS, Hojat M. Relationship between performance in medical school and postgraduate competence. *J Med Educ* 1983; 58: 679–685.
34. Stimmel B, Serber M. The role of curriculum in influencing students to select generalist training: a 21-year longitudinal study. *J Urban Health* 1999; 76: 117–126.
35. Ferrier BM, Woodward CA. Does premedical academic background influence medical graduates' perceptions of their medical school or their subsequent career paths and decisions? *Med Educ* 1983; 17: 72–78.
36. Dickman RL, Sarnacki RE, Schimpfhauser FT, Katz LA. Medical students from natural science and nonscience undergraduate backgrounds. Similar academic performance and residency selection. *JAMA* 1980; 243: 2506–2509.
37. Sierles FS, Vergare MJ, Hojat M, Gonnella JS. Academic performance of psychiatrists compared to other specialists before, during, and after medical school. *Am J Psychiatry* 2004; 161: 1477–1482.
38. Fishbein RH. Origins of modern premedical education. *Acad Med* 2001; 76: 425–429.
39. The GPEP report: report of the panel on the general professional education of the physician and college preparation for medicine. Available at: http://www.eric.ed.gov/ericdocs/data/ericdocs2sql/content.storage_01/0000019b/80/2e/c3/56.pdf. Accessed May 2009.
40. Dienstag JL. Relevance and rigor in premedical education. *N Engl J Med* 2008; 359: 221–224.
41. Doblin B, Korenman S. The role of natural science in the premedical curriculum. *Acad Med* 1992; 67: 539–541.
42. Brieger GH. The plight of premedical education: myths and misperceptions—part II: science “versus” the liberal arts. *Acad Med* 1999; 74: 1217–1221.
43. Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner report. *N Engl J Med* 2006; 355: 1339–1344.
44. Hoover EL. A century after Flexner: the need for reform in medical education from college and medical school through residency training. *J Natl Med Assoc* 2005; 97: 1232–1239.
45. Gross JP, Mommaerts CD, Earl D, De Vries RG. Perspective: after a century of criticizing premedical education, are we missing the point? *Acad Med* 2008; 83: 516–520.
46. Montgomery K, Chambers T, Reifler DR. Humanities education at Northwestern University's Feinberg School of Medicine. *Acad Med* 2003; 78: 958–962.
47. Sirridge M, Welch K. The program in medical humanities at the University of Missouri-Kansas City School of Medicine. *Acad Med* 2003; 78: 973–976.
48. Krackov SK, Levin RI, Catanese V, et al. Medical humanities at New York University School of Medicine: an array of rich programs in diverse settings. *Acad Med* 2003; 78: 977–978.
49. Kuper A. Literature and medicine: a problem of assessment. *Acad Med* 2006; 81(suppl): S128–S137.
50. Spike JP. Developing a medical humanities concentration in the medical curriculum at the University of Rochester School of Medicine and Dentistry, Rochester, New York, USA. *Acad Med* 2003; 78: 983–986.
51. Murray J. Development of a medical humanities program at Dalhousie University Faculty of Medicine, Nova Scotia, Canada, 1992–2003. *Acad Med* 2003; 78: 1020–1023.
52. Dittrich LR. Preface. *Acad Med* 2003; 78: 951–952.
53. Feigelson S, Muller D. “Writing about medicine”: an exercise in reflection at Mount Sinai. *Mt Sinai J Med* 2005; 72: 322–326.
54. Bardes CL, Gillers D, Herman AE. Learning to look: developing clinical observational skills at an art museum. *Med Educ* 2001; 35: 1157–1161.
55. Charon R, Banks JT, Connelly JE, et al. Literature and medicine: contributions to clinical practice. *Ann Intern Med* 1995; 122: 599–606.
56. Karkabi K, Ungar L, Kaffman M, et al. The encounter between literature and medicine— from theory to practice [in Hebrew]. *Harefuah* 2008; 147: 350–353, 372.

57. American Medical Association. Initiative to transform medical education. Available at: <http://www.ama-assn.org/ama1/pub/upload/mm/377/finalitme.pdf>. Pg 3 Accessed May 2009.
58. Gold AP. Nurturing humanism in medicine. HYPERLINK "<http://www.shma.com/>" *Sb'ma: A Journal of Jewish Responsibility*. March 2002 (Volume 33: 599).
59. Cohn F, Lie D. Mediating the gap between the white coat ceremony and the ethics and professionalism curriculum. *Acad Med* 2002; 77: 1168.
60. Wachtler C, Lundin S, Troein M. Humanities for medical students? A qualitative study of a medical humanities curriculum in a medical school program. *BMC Med Educ* 2006; 6: 16.
61. Shapiro J, Rucker L. Can poetry make better doctors? Teaching the humanities and arts to medical students and residents at the University of California, Irvine, College of Medicine. *Acad Med* 2003; 78: 953–957.
62. Cohen PA, Fortin AH VI. Curriculum of literature and medicine for residents. *Acad Med* 1999; 74: 578–579.
63. Chen D, Lew R, Hershman W, Orlander J. A cross-sectional measurement of medical student empathy. *J Gen Intern Med* 2007; 22: 1434–1438.
64. DiLalla LF, Hull SK, Dorsey JK. Effect of gender, age, and relevant course work on attitudes toward empathy, patient spirituality, and physician wellness. *Teach Learn Med* 2004; 16: 165–170.
65. Hojat M, Mangione S, Nasca TJ, et al. An empirical study of decline in empathy in medical school. *Med Educ* 2004; 38: 934–941.
66. Hojat M, Gonnella JS, Mangione S, et al. Empathy in medical students as related to academic performance, clinical competence and gender. *Med Educ* 2002; 36: 522–527.
67. Mangione S, Kane GC, Caruso JW, et al. Assessment of empathy in different years of internal medicine training. *Med Teach* 2002; 24: 370–373.
68. West CP, Huntington JL, Huschka MM, et al. A prospective study of the relationship between medical knowledge and professionalism among internal medicine residents. *Acad Med* 2007; 82: 587–592.
69. DasGupta S, Charon R. Personal illness narratives: using reflective writing to teach empathy. *Acad Med* 2004; 79: 351–356.
70. Shapiro J, Morrison E, Boker J. Teaching empathy to first year medical students: evaluation of an elective literature and medicine course. *Educ Health (Abingdon)* 2004; 17: 73–84.
71. Veloski JJ, Fields SK, Boex JR, Blank LL. Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. *Acad Med* 2005; 80: 366–370.
72. DasGupta S, Meyer D, Calero-Breckheimer A, et al. Teaching cultural competency through narrative medicine: intersections of classroom and community. *Teach Learn Med* 2006; 18: 14–17.
73. Stern DT, Frohna AZ, Gruppen LD. The prediction of professional behaviour. *Med Educ* 2005; 39: 75–82.
74. Reed DA, West CP, Mueller PS, et al. Behaviors of highly professional resident physicians. *JAMA* 2008; 300: 1326–1333.
75. Goldberg JL. Humanism or professionalism? The white coat ceremony and medical education. *Acad Med* 2008; 83: 715–722.
76. Ball S, Bax A. Self-care in medical education: effectiveness of health-habits interventions for first-year medical students. *Acad Med* 2002; 77: 911–917.
77. Austenfeld JL, Paolo AM, Stanton AL. Effects of writing about emotions versus goals on psychological and physical health among third-year medical students. *J Pers* 2006; 74: 267–286.
78. Lypson ML, Hauser JM. Talking medicine: a course in medical humanism— what do third-year medical students think? *Acad Med* 2002; 77: 1169–1170.
79. Kumagai AK. A conceptual framework for the use of illness narratives in medical education. *Acad Med* 2008; 83: 653–658.
80. Reed VA, Jernstedt GC, Boudreau D, et al. Meeting the challenge of MSOP: comprehensive measurement of profiles of student characteristics across medical schools. *Acad Med* 1999; 74(suppl): S96–S98.
81. Stratton TD, Elam CL, Mcgrath MG. A Liberal Arts Education As Preparation For Medical School: How Is It Valued? How Do Graduates Perform? *Acad Med*. 2003 Oct; 78(10 Suppl): S59–61.

Mission-Based Budgeting for Education: Ready for Prime Time?

Michelle Sainté, BS,¹ Steven L. Kanter, MD,² and David Muller, MD¹

¹Department of Medical Education, Mount Sinai School of Medicine, New York, NY

²University of Pittsburgh School of Medicine, Pittsburgh, PA

ABSTRACT

Mission-based budgeting has been touted as the preferred method of assessing and assigning departmental budgets in academic medical centers. Mission-based budgeting in its simplest form is a methodology that allows an institution's finance department to align costs with actual activities (typically clinical care, administration, research, and teaching). Despite its intuitive appeal, a minority of the academic medical centers across the country have embraced it. Mount Sinai School of Medicine was among the first to incorporate this approach, and we have almost a decade of experience with its risks and benefits. This article focuses on the education component of mission-based budgeting. We review all aspects of mission-based budgeting, including (1) the many variables that must be factored into a metric and how those variables differ among institutions, (2) the metric itself and how it reflects institutional philosophy, (3) the impact of mission-based budgeting on faculty and chairs of departments, (4) existing processes that ensure the quality, reliability, and validity of the different mission-based budgeting systems, and (5) a comparative perspective. *Mt Sinai J Med* 76:381–386, 2009. © 2009 Mount Sinai School of Medicine

Key Words: mission-based budget, paying for education, teaching metric.

In the early 1990s, Dr. Robert Watson and his colleagues¹ at the University of Florida College of Medicine began thinking of ways to allocate

resources based on institutional principles. The changing climate of healthcare delivery and financing in the mid-1990s had a major impact on the process used by medical schools to fund academic programs and support education. Historically, the budget process at most academic medical centers was initiated and controlled on a departmental level. Chairs and institute/center directors created budgets based on previous spending patterns, with incremental increases for strategic growth and planning, and subsequently negotiated with deans for these resources. In many cases, these processes included an element of arbitrariness and did not hold departments and faculty accountable for the department's or institution's ability to fulfill their missions. This process was also devoid of a system designed to account for physicians' time and effort.

During this same time frame, the research superstructures (eg, the National Institutes of Health and Howard Hughes Medical Institute) began to require faculty investigators to include their effort when submitting grant budgets, and institutions began to respond proactively and require that faculty effort be documented and itemized. This movement encouraged academic medical centers to create, on an institutional level, systems to document faculty effort related to the research enterprise.

In 1998, a national panel was appointed by the Association of American Medical Colleges to create an infrastructure for institutions to understand and identify categories for faculty effort and the variables associated with each aspect of their missions [mission-based budgeting (MBB)].² In this system, 4 areas were outlined for grouping faculty effort in education: direct teaching, creation of educational products, education administration or service, and scholarship. These categories allowed for flexibility within the model and created a foundation for schools to recognize a wide array of faculty efforts. Once educational activities were defined, the authors created a stepwise system to develop a relative value metric. The first step was to identify the activity in the

Address Correspondence to:

Michelle Sainté

Department of
Medical Education
Mount Sinai School of Medicine
New York, NY
Email: michelle.sainte@mssm.edu

context of the aforementioned groups. The second step involved understanding the performance from a quantitative perspective. In rare cases, this step was expanded to include quality as well. The third step was to assess the institutional value of each role in achieving the educational mission. The final step was to adjust for the program in which the educational activity was being performed, as some institutions valued work done in the required curriculum at a higher level than work done in ancillary or elective programs. This model was designed to encourage and support institutions interested in creating an MBB system. It also served as a mechanism to evaluate faculty effort and as a foundation for reward systems for extraordinary faculty accomplishments. The seminal article in the area of MBB was published in 1999 and described the University of Florida model in detail.¹ That system comprised a multilayered approach to funding the mission of education. The scope was broader than most and included medical and graduate programs, graduate medical education, committee work, and education. This was the first report on a new paradigm for allocating resources dedicated to teaching.

At Mount Sinai as well as many other academic medical centers, the research and clinical enterprises were the first and least complex to create a system to document effort and identify parameters for evaluating excellence. Identifying research density and tabulating effort on grants received facilitated the understanding of the faculty effort in the research mission. Performance measures such as visit volume data, earned relative volume units, and mean denial rates were tools used to document faculty effort in clinical operations. Education was a more difficult mission to evaluate. Despite great strides, the assessment of quality and scope related to education continues to lag far behind analogous metrics for research and clinical care.

NATIONAL PERSPECTIVE

We searched the literature and engaged in conversations with faculty and staff at several academic medical centers, both private and public. When we looked at the national experience in MBB for education, we learned that (1) there are a great number of variables involved in creating a metric and (2) there is a very wide range of experiences among medical schools.

Some of the variables that go into creating a metric include the following:

- Actual dollar amount per contact hour of teaching.
- Amount of prep time allotted for each contact hour.

- Distinctions between types of teaching contact (lecture, laboratory, review sessions, etc.).
- Definition and reimbursement of leadership roles in education (eg, course directors).
- Definition and reimbursement of the administrative costs of directing courses/clerkships.
- Definition of covered education activities (teaching, mentoring, and committee work).
- Factoring in quality.

The spectrum of MBB experiences among medical schools includes the following:

- Schools that have neither planned nor implemented.
- Schools that have planned and never implemented.
- Schools that have planned and implemented.
- Schools that have planned, implemented, and failed to sustain.
- Schools that have planned, implemented, and evolved over time.

In looking at the national experience in MBB for education, the published literature predominantly describes processes that have been developed and implemented in public schools of medicine. More important to the process than the actual variables mentioned previously are the philosophical underpinnings of MBB and the ways in which these variables are related to an institution's philosophy. One approach is to micromanage the variables in an effort to perfect a system of value-based compensation for education. Another is to appreciate that the entire educational enterprise must fit within the financial constraints of resources that are available to pay for education. Ultimately, many educators believe that the specific value of each of the variables that make up the metric is not nearly as important as making sure that education is recognized and rewarded.

At most institutions, including Mount Sinai, the impetus for implementation began with a primary goal of reallocating resources in a time of fiscal uncertainty. When we began our research, we contacted over 10 schools who had documented their MBB efforts. We searched the literature for standard terms used in this process: MBB and management, resource allocation, and faculty effort. There were several institutions that had gone through the painstaking process of working with teams of faculty to create complex MBB systems.¹ After contacting several schools documented in this article, we found that many of the cited schools never implemented their systems. There has been mention of proposed MBB systems in a handful of private schools; however, after combing the literature, we

did not find any detailed descriptions published of programs that have actually been implemented.¹⁻³ In this article, we review MBB as it has been implemented in a variety of public schools of medicine, describe the method that we use at Mount Sinai in detail, and reveal common elements that can be considered best practice regardless of the type of school.

In the cited articles, the majority focused on a single institution's model and the methods used to implement it.^{1,4,5,6,7,8} In 1998, Creighton School of Medicine initiated an MBB strategy as a result of its selection by the Association of American Medical Colleges to participate in a pilot project.⁵ The article describing its efforts focuses on the fiscal stability of its Department of Internal Medicine. The Creighton process began with a concern related to the huge losses suffered by that department. This issue was directly attributed to faculty effort and the changing needs of the educational program. Aligning the department's resources with its mission seemed to be a logical approach to minimizing negative perceptions and enhancing the faculty's commitment to the diversity of needs across divisions. The goals for initiating this project included identifying objectives for the academic missions, enhancing the productivity of faculty and departments with respect to these objectives, advancing the financial achievements of departments, encouraging accountability, and supporting collaboration. The project resulted in 4 innovations: the creation of faculty activity profiles, a tool to measure faculty performance, an MBB model, and a revision of the leadership model within the school.

In 2002, Mallon and Jones² reviewed 41 schools and/or departments, including the University of Florida, that had developed metrics to quantify the education productivity of faculty. The authors assembled a database of faculty educational productivity and analyzed the reasons that each metric was chosen, what it specifically assessed or tracked, and how it reflected the information collected. This article focused on faculty effort independent of the allocation of fiscal resources. Because of this focus on effort without regard for available resources, faculty and school leadership have the creativity to define what, if any, modifications should be made to how faculty spend their time and to categorize effort based on the institution's values, not based on what it can afford. Although several challenges were identified in the application of these measurement systems, the authors concluded that quantifying the value of the educational mission and identifying appropriate support could be achieved with this type of approach.

The Dalhousie University Faculty of Medicine has the oldest system in the literature, which was documented in 2003 as a retrospective review.⁵ Its process, implemented in 1993, allocated funds not on the basis of on a metric but on the basis of defined departmental educational goals and attributed a specific portion of the budget to each goal. The authors suggested that this type of system is successful only when the educational outcomes, not specific teaching activity, are used to measure success. This system was met with resistance, and many faculty were unclear on the parameters for success. It was critically important to align the finances to the mission and equally important to offer the faculty transparency of goals and educational outcomes.

The University of Pennsylvania model,⁶ published in 2007, identifies clearly defined parameters used for budgeting. Its method includes transparency of process and distinct areas of allocation for teaching, research, and administrative duties for budget flow. The system was initiated in 2006. The Penn model is used to quantify effort and to support didactic and small-group medical student teaching, supervision of residents, and administrative support of educational programs. However, the primary focus of the Penn process was to ensure that an appropriate balance in funding for the clinical operations was achieved, and the model was developed for clinical departments only. The Penn approach had a positive overall impact on departments that had disproportionate teaching loads and equalized the distribution of funds to those departments for which education and research were underreported.

A web-based MBB was reported in 2002 by faculty at the University of California Davis campus.⁷ This process was in response to a need to define productivity, ensure faculty assessment, and understand program management. This system was created to identify qualitative and quantitative measures of faculty effort. The article identified several barriers to success, including faculty resistance and cultural norms that were in direct conflict with the MBB ideology. The final system was implemented in 2002 but was used merely as a reporting structure, not as a management or budgeting tool. The resistance from faculty and the lack of literature demonstrating success at other institutions, at that time, decreased the momentum for the creators. Their system continues to be used for career planning and development and as a platform for discussions related to performance appraisal and goal setting.

The first fully implemented mission-based model published in the literature that includes a quality component was described by Stagnaro-Green *et al.*

in 2008.⁸ In the 2004–2005 academic year, the New Jersey Medical School included a quality factor as part of its mission-based funding mechanism: 80% of the education budget was used to remunerate departments for their teaching effort, with the remaining 20% apportioned on the basis of quality. The quality allocation was quantified with student and faculty peer evaluations of courses and clerkships, faculty compliance in administrative duties, and faculty development and scholarship. The overall quality ranking was based on a gestalt of all areas reviewed, not a dedicated metric. The system described financial rewards to departments for the quantity and quality of their faculty's educational programs. This method does not account for individual faculty effort and on the basis of the description does not require that faculty document each hour dedicated to their teaching or leadership in the educational program. According to the authors, these omissions have increased the buy-in of faculty and chairs and has made implementation of the quality component much more palatable.

COMPARATIVE PERSPECTIVE

In 2002, Mount Sinai School of Medicine (MSSM), in the midst of financial challenges, recruited a consulting firm to develop a more efficient, effective, and transparent budget process that reflected the institution's core missions and linked resources to those missions. As part of a reengineering process largely driven by external consultants, the School of Medicine implemented an MBB system that made allocations in 5 categories: clinical, administrative, research, teaching, and strategic (CARTS).

The Department of Medical Education at Mount Sinai was responsible for researching, designing, and implementing a system to calculate teaching costs associated with each student contact hour across all years of medical and graduate school education. The result was a teaching dollar (T dollar) database for the medical and graduate schools that allowed the school to formally acknowledge individual faculty service to the institution and provide compensation to departments in direct proportion to faculty teaching effort. Important factors that were considered included setting a dollar amount per contact hour, associated prep hours per contact hour, defining and compensating leadership roles, establishing the cost of administrative support per course, and defining the weighted metric between lecture, small-group, laboratory, and clinical (bedside) teaching.

The initial T dollar system paid a fixed dollar amount per contact hour and added a factor of prep

hours for all teaching contact hours regardless of venue (lecture, laboratory, or small group) in the first 2 years. Clinical clerkship teaching was recognized as being more fluid, with faculty preceptors and students assigned to inpatient teams and interacting throughout the day. As a result, departments were paid a fixed dollar amount depending on the length of the rotation (number of weeks) and how many students participated in any given clerkship. All course and clerkship directors were paid according to the same fixed T dollar educational leadership metric.

The University of Pittsburgh School of Medicine (UPSOM) implemented in 1997 an educational credit unit (ECU) system that shares many of the features of the MSSM system. However, instead of the payment of a fixed dollar amount per ECU (or per contact hour), a fixed sum of money was distributed to a department according to the percentage of ECUs generated collectively by faculty members with a primary appointment in that department. This has the advantage of working with a fixed dollar amount, which is known in advance, and simplifies the budgeting process. However, disadvantages of the ECU system include that the value of a single ECU decreases as new educational activities are covered by the system.

At MSSM, the data used to determine reimbursement for the teaching effort report are compiled from the curriculum database and are a reasonably accurate account of the time and effort involved in the educational program during the previous academic year. The database is maintained by the Office of Curriculum Support and is updated at the end of each term. These data are submitted to the Office of Academic Administration and translated into a set of financial data that are then sent to each department for review and confirmation. Once the departmental review is completed, a final report is sent to the Dean's Office, and funds are distributed as part of the departmental budget disbursement. Each department is privy to detailed information on the annual teaching activity of its faculty, and department chairs employ various degrees of transparency in associating individual salaries with the T dollar report. For example, some chairs provide faculty with individualized salary breakdowns based on the CARTS methodology, whereas others simply lump T dollars into the whole of the departmental CARTS budget and pay faculty without providing a detailed report of the actual sources of support for that salary. UPSOM uses a similar approach. Data from course and clerkship syllabi, as well as other sources, are collected by staff of the Office of Medical Education and used to seed a database. Each individual faculty member is asked

to confirm or modify his or her data. The staff of the Office of Medical Education then run a series of validity checks, calculate totals for faculty members and departments, and deliver to the financial management team the percentage of ECUs generated by each department.

Many nuances presented themselves in MSSM's early years of MBB and were addressed in an ad hoc fashion. Examples included whether to provide T dollars for review sessions, office hours, and examinations, whether co-associate or associate directors had legitimate leadership roles that should be formally compensated, and whether T dollars should be provided for elective experiences. At UPSOM, early dilemmas included whether to include committee service related to the educational program, whether to include teaching students not enrolled in the medical school, and whether to credit faculty with volunteer or adjunct appointments.

In 2006, we implemented a continuous quality improvement initiative and reviewed the process and metrics used since 2002. The most important outcome of this initiative was the understanding that the cost of education could not increase in an unrestricted manner. Before 2006, there was a fixed dollar amount per faculty contact hour/week regardless of course length or the number of faculty. For example, a course director could unilaterally choose to increase the number of faculty per small group from 2 to 3, in effect increasing the cost of that exercise by 50%. After 2006, we recognized that the pool of resources available for teaching should be finite and such an increase in faculty, if deemed pedagogically appropriate, would translate into a lower dollar amount per faculty contact hour or week.

This and some minor enhancements to the MBB system were developed and implemented by faculty and chair stakeholders working closely with the Department of Medical Education and were communicated broadly to faculty in an effort to maintain complete transparency.

MSSM is currently engaged in a second quality review that, when completed, will incorporate measures of teaching quality and administrative excellence into T dollar disbursement. An ad hoc working group of educators and administrators has identified several quality factors that will allow individual faculty members to be rewarded for excellence in teaching. These include metrics directly related to teaching (eg, student, self, and peer evaluations, quality of web-based resources, and quality of faculty development) and adherence to administrative responsibilities (eg, timely posting of schedules, syllabi, and grades, clear goals and

objectives, and providing evaluations and feedback to faculty).

Once the quality measures have been determined, a proposal will be presented to an advisory group of chairs and senior administrators for feedback. Our intention is to reward faculty for outstanding performance by enhancing their T dollar disbursement and incentivize others to improve their teaching and compliance with administrative duties.

CONCLUSION

Most of the institutions that currently have an MBB program implemented their structure between 2000 and 2003. The spectrum of experience is vast, and there are still many schools that have not, and do not plan to, implement a system of their own. It appears that the most highly developed systems are those that define education broadly (ie, they include activities such as elective teaching, mentorship, and committee work) and have a flexible structure that can be adjusted to meet their educational needs and fiscal limitations.^{2,4,5,7,8}

In our institution, the most successful application of MBB includes transparency of process and resource allocation and periodic reviews by stakeholders. At Mount Sinai and around the country, MBB has led to a major shift in the ways in which chairs, deans, and faculty think about and value education. Because education can now be quantified and rewarded, medical schools have taken a major step toward closing the historical priority gap between research and clinical care on the one hand and teaching on the other hand. In our experience, faculty who teach are now able to justify their time away from patients or their labs and chairs are more eager to encourage interested and talented faculty to pursue academic careers as educators.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Watson RT, Romrell LJ. Mission-based budgeting: removing a graveyard. *Acad Med* 1999; 74: 627–640.
2. Mallon W, Jones R. How do medical schools use measurement systems to track faculty activity and productivity in teaching. *Acad Med* 2002; 77: 115–123.

3. Nutter D, Bond J, Collier B, et al. Measuring faculty effort and contributions in medical education. *Acad Med* 2000; 75: 200–207.
4. Rich E. Departmental and school-wide mission-based management: a case study. *Am J Med* 2001; 110: 245–251.
5. Reudy J, MacDonald N, MacDougall B. Ten-year experience with mission-based budgeting in the faculty of medicine of Dalhousie University. *Acad Med* 2003; 78: 1121–1129.
6. Kennedy D, Johnston E, Arnold E. Aligning academic and clinical missions through an integrated funds-flow allocation process. *Acad Med* 2007; 82: 1172–1177.
7. Howell L, Hogarth M, Anders T. Creating a mission-based reporting system at an academic health center. *Acad Med* 2002; 77: 130–138.
8. Stagnaro-Green A, Roe D, Soto-Green M. Development of a mission-based funding model for undergraduate medical education: incorporation of quality. *Teach Learn Med* 2008; 20: 168–173.

INSPIRE: The Individual Scholarly Project and Independent Research Experience in Undergraduate Medical Education

Karen Zier, PhD, and Lisa D. Coplit, MD

Department of Medical Education, Mount Sinai School of Medicine, New York, NY

ABSTRACT

Individualized learning is a fundamental tenet of the Carnegie Foundation's new recommendations for physician training. A primary goal of Mount Sinai School of Medicine's new curriculum is to train self-directed physicians who have mastered lifelong learning skills. The Individual Scholarly Project and Independent Research Experience (INSPIRE) was created to enable fourth-year students to conduct mentored, independent scholarly projects to develop critical thinking skills and intellectual independence. Four students were accepted into the 2008 12-week pilot. Two did clinical research, 1 conducted medical education research, and 1 continued a basic science project. INSPIRE featured weekly sessions in which students shared their progress, heard about the careers of physician-scientists, and participated in presentation skills workshops. Mentors gave feedback using a form based on the program goals. Anonymous evaluations showed that all participants believed their goals were successfully fulfilled, that the likelihood of including research in their careers increased, and that they felt more skilled at writing abstracts and presenting their work orally. Students agreed that INSPIRE was a valuable opportunity for acquiring in-depth knowledge on a topic and building research and presentation skills. Helping students shape an individualized scholarly experience may help to produce doctors who are self-directed, are personally fulfilled, are able to serve society with diverse expertise, and have the tools to become leaders

in their fields. *Mt Sinai J Med* 76:387–391, 2009. © 2009 Mount Sinai School of Medicine

Key Words: individualized learning, independent research, INSPIRE, undergraduate medical education.

Individualized learning, integration of knowledge, identity formation, and the pursuit of excellence, including advancing the field of medicine through scholarship, are key to the Carnegie Foundation's new recommendations for physician training.¹ Consistent with these goals, the Association of American Medical Colleges' Task Force II on Clinical Research has called for medical schools to educate students in the basic principles of clinical-translational research, develop their ability to critique published studies, and enable them to participate actively in studies that would nurture their interest in clinical research.² Successful fulfillment of these goals would help students to sharpen their critical thinking skills and express their individual creativity and might encourage students to pursue careers as physician-scholars.

Many leading medical schools devote curricular time to required or elective scholarly experiences, although the time available varies greatly. Most of these schools require that students complete some coursework and that they produce a scholarly project, such as a written thesis, poster, or oral talk based on their projects. Schools including Yale, the University of Pittsburgh, Stanford, and Duke are among those that require students to conduct a scholarly project.^{3–6} In contrast, the University of California San Francisco, Brown, and the University of Pennsylvania offer students the choice of an elective scholarly program.^{7–9} At each of the schools, students can complete their projects within 4 years of school, although many elect to devote an entire year to their projects in order to be able to perform a more in-depth study, thus adding a year to their medical education.

Address Correspondence to:

Karen Zier, PhD

Department of Medical
Education, Mount Sinai School
of Medicine New York, NY
Email: karen.zier@mssm.edu

MEDICAL STUDENT RESEARCH AT MOUNT SINAI

Medical students at Mount Sinai have long demonstrated interest in scholarship.¹⁰ Participation in the Summer Research Scholars Program, which supports clinical, translational, basic, outcome, community-based, and global health projects, has ranged from 56 for the class of 2004 to 100 for the class of 2011. In addition, the number of students choosing to do a scholarly year, a year-long pullout program most often supported by foundations including the Howard Hughes Medical Institute and the Doris Duke Charitable Foundation or the National Institutes of Health, has increased from 5 in 1997 to 30 in 2008. Students are encouraged to present the results of their scholarship at the annual Medical Student Research Day, and participation is very strong, with the number of student participants ranging from 19 in 1997 to 115 in 2008. In the Mount Sinai School of Medicine (MSSM) 2007 graduation survey, of the 80% of respondents who participated in research during medical school, 67% recommended that students pursue research during medical school. Fifty-two percent reported a positive impact of their research on their career, and 86% reported that their research experience had enhanced their understanding of the role of research in patient management. These data suggest that students value their research experiences and that they affect their clinical careers.

PLANNING THE INDIVIDUAL SCHOLARLY PROJECT AND INDEPENDENT RESEARCH EXPERIENCE (*INSPIRE*)

MSSM is currently undergoing an exciting curriculum reform with the goal of producing physicians who are capable of promoting change as advocates, clinicians, and scientists (Table 1). As one of the first efforts toward achieving these goals, the core curriculum planning group decided to introduce a substantial scholarly component into the undergraduate medical curriculum: *INSPIRE*.

INSPIRE was designed to provide students with protected time to conduct a strongly mentored project that would further develop critical thinking skills, inspire creativity, and promote intellectual independence. It was less important what type of scholarship students pursued and more important that they learn the value of scholarship within the study of medicine. In fact, students would be encouraged to

Table 1. *Goals of the New Mount Sinai School of Medicine Curriculum.*

-
- Integration across scientific disciplines
 - Integration between basic science and clinical medicine
 - Longitudinal clinical experiences as a focus of clinical training
 - An experience that is broader and deeper than traditional models of medical education
 - An educational paradigm that promotes self-directed and lifelong learning
-

be true to their interests and creative in their project choices. In 2007, as the first step in planning, a development committee composed of experts in the various fields in which students might pursue their scholarly work was convened. Committee members represented clinical/translational research, advocacy and community service, community medicine/health policy, medical humanities, global health, and medical education. The committee also included students from the second- and fourth-year classes who provided insight into what students would like to have in a scholarly program or would have liked to have had in a scholarly program, respectively. The goals of *INSPIRE* were structured to contribute to the mission to provide individualized scholarly immersion experiences for medical students within a 4-year curriculum; to provide unique options for in-depth study, including nontraditional scholarly experiences that would enrich students' future practice of medicine; and to attract students to MSSM who are self-directed and interested in a school that offers students a chance to have a scholarly experience. With these basics, this group took on the charge of conceptualizing *INSPIRE* from a broader perspective without considering the boundaries of the current curriculum.

THE *INSPIRE* PILOT

Initially, an almost year-long experience was envisioned. However, freeing such a large period of time for individualized study will depend on completing the curriculum reform. Thus, the committee planned a pilot program that could be inserted into the existing curriculum while maintaining the goals of *INSPIRE*. The *INSPIRE* pilot was created as a 12-week curriculum (September to December 2008) during year 4 that included a mentored scholarly project and a didactic curriculum (Table 2)

Although it was anticipated that projects would concentrate in areas of institutional strength, rigorous scholarship outside traditional research areas was

Table 2. *INSPIRE Didactic Curriculum.*

-
- Introductory Meeting of Participants and Discussion of Projects
 - Preparing Your Talk and Written Proposal
 - How to Give a Talk
 - Updates (brief, informal presentations on progress and problems)
 - Writing an Abstract that Delivers Your Message
 - Presentation Skills Workshop– Part 1
 - Presentation Skills Workshop– Part 2
 - Updates
 - Preparing a Successful Poster
 - Mount Sinai Medical Student Research Day
 - Dress Rehearsal for INSPIRE Research Day (videotaped)
 - Presentation Skills Workshop– Part 3 (debriefing of videotapes)
 - INSPIRE Research Day
-

encouraged. Students could choose to do an in-depth study in their chosen specialty or to explore a related area that they believed would enrich their practice of medicine. A call for applications was sent to the third-year class, and an information session was held at a class meeting. The formal application required a project proposal, which was prepared with the student's mentor, and a mentor's letter of support. Students were able to apply with carefully planned new projects or to continue projects begun earlier in their training. Only students with no academic difficulties were eligible for the pilot.

From extensive experience in creating medical student research experiences at MSSM, it seemed clear that excellent mentorship and a project with realistic goals would be keys to the students' success. Students interested in applying to INSPIRE met with the program directors months or weeks before the due date to get help in identifying outstanding potential mentors and to discuss the feasibility of their proposed projects. One program director, who is a PhD, is a translational scientist and medical educator, and the other, who is an MD, is a clinician-educator. Thus, program advisors were available who were able to discuss the various types of research projects that students were considering. INSPIRE took place early

in the academic year. This scheduling undoubtedly reduced the potential number of applications because participation meant that there would be little time to do electives at other institutions before residency applications were due. Moreover, with the current required clerkships in the fourth year and time needed for interviews, students who participated in INSPIRE were unable to accommodate many additional clinical electives. Four students applied, and all were accepted. This small first class had the advantage of permitting close monitoring of the new program. Each application proposed a project that was rigorous, was feasible for completion within 12 weeks, and asked an important, novel, and ethical question. Three students began new projects, and 1 continued a project that she had begun during a scholarly year (Table 3).

Two projects were in clinical research, 1 was in basic science, and 1 was in medical education. An interactive curriculum was created to introduce participants to the career path of the physician-scholar and teach important writing and presentation skills (Table 2). Weekly 1-hour meetings provided students with support from the program directors, allowed students to establish a group identity, enabled them to meet faculty members who were clinician-scholars and role models, and gave them the opportunity to present progress reports on their projects. Students received in-depth feedback at these sessions on abstracts and posters that they had previously prepared for Medical Student Research Day and on the video-taped oral presentations given during INSPIRE. In addition, program mentors gave students structured feedback at the midway point and at the end, using a standard form based on the goals of the program. At the conclusion of the program, students presented their work as 20-minute oral talks on INSPIRE Research Day. These final presentations were evaluated by 2 experienced faculty research members not involved with INSPIRE who used criteria taught during their presentation workshops. Students were graded with a pass/fail system, and they earned 12 weeks toward their elective credit requirement.

Table 3. *INSPIRE Projects and Supporting Departments.*

-
1. Secondary Hyperparathyroidism in Patients Taking Tenofovir (Department of Medicine/Division of Liver Disease)
 2. Stress Testing and Biomarkers for the Evaluation of Intermediate Risk ACS (Department of Medicine/Division of Cardiology)
 3. Evaluating the Merits of Course Review: A Qualitative Analysis of 20 Educators' Experiences (Department of Medical Education)
 4. Project FAT10: A Key Mediator of HIV Viral Protein R in HIV-Associated Nephropathy (Department of Medicine/Division of Nephrology)
-

Abbreviation: ACS, acute coronary disease; HIV, human immunodeficiency virus

EVALUATION OF THE 2008 *INSPIRE* PILOT

Students completed an anonymous online evaluation of the program. Evaluations of the mentors were treated as confidential but could not be anonymous because only 1 student worked with each mentor. A 1-hour session after the program ended was dedicated to receiving in-person feedback from the students to improve the program for future participants. Because only 4 students participated in the pilot program, the quantitative data are limited, and the feedback obtained is preliminary. However, the evaluations are useful as they provide direction to the program leadership about how to better meet the students' needs.

All 4 students assessed their *INSPIRE* experience as excellent. Three felt that their goals for *INSPIRE* were fulfilled very successfully, and 1 felt that they were fulfilled successfully (Table 4).

Two students strongly agreed and 2 agreed that participation in *INSPIRE* increased the likelihood that they would include a research component in their careers. All 4 students agreed that they were better

able to write an abstract describing the results of a research study. Three strongly agreed and 1 agreed that they were better prepared to give an oral research presentation. Responses to open-ended questions revealed that the students felt *INSPIRE* was valuable because it offered them the opportunity to carry out an independent scholarly project toward the end of their medical education after they had exposure to both basic science courses and clinical clerkships, because they had the chance to acquire in-depth knowledge in a focused area, because they could benefit from sharing the progress and experiences of their peers in the program, and because it provided a chance to work closely with a mentor (Table 5). In the evaluation and in an in-person debriefing session, students were asked to provide suggestions to improve the program and to describe the impact of the experience on their current and future careers. Suggestions for improvement were consistent and included increasing the length of the program, adding didactic sessions on designing a research study, and providing more statistical support to help with data analysis. In 2009, on the basis of student feedback, a module on research methodology will be incorporated into the didactic curriculum.

Table 4. *Evaluation of INSPIRE by Participants.*

	Strongly Agree	Agree	Disagree	Strongly Disagree
Goals for <i>INSPIRE</i> were fulfilled	3	1	0	0
Increased likelihood of a research component in career	2	2	0	0
Better able to write an abstract on research	4	0	0	0
Better able to give an oral research talk	3	1	0	0

Table 5. *Student Comments: The Value of INSPIRE for Future Applicants.*

- [It] provides time to pursue scholarly work outside of the usual research elective.
- This program is very important for people in [their] fourth year who have begun to develop specific interests in a field and want to get a better sense of what it means to do research/how important it will be to them to incorporate this into their career. Selecting a residency program is the first of many forks in the road, and it's timely for students to learn something about their own priorities before going through this process. Furthermore, there's definite value in acquiring in-depth knowledge on a topic of interest (both personally and professionally) before plunging into residency. Since not everyone will take a year off, it's very important for people to have the option to explore this at a more "mature" phase of their training, as opposed to just the summer after first year.
- *INSPIRE* in its current form teaches students how to present existing data, in the form of abstract and oral presentation, with an emphasis on the latter. It fosters helpful discussion about diverse projects that the other students are conducting. *INSPIRE* gives a brief but important glimpse into the realities— including the rewards and challenges (hard work but slow progress)— of conducting research, whether clinical or laboratory-based.
- I think *INSPIRE* is really valuable for continuing a project in an area of interest or for creating a brief clinical or educational research study, and it was of particular value for me since this was my first real research project. I felt like that was definitely a missing piece to my medical school experience, and this allowed me the opportunity to design and pursue my own research without having to give up a whole year to do so. The didactic curriculum is of particular value to hone presentation and teaching skills, which is definitely not found in other parts of the general curriculum. I think it's also really valuable to help cultivate mentorships during application season— a very formative time when personal career advice is always welcome.

For 2009, 8 students have been accepted. As in 2008, several more students were interested in participating, but current limitations on the ability to do clinical electives, as discussed previously, prevented this.

CONCLUSIONS AND FUTURE PLANS

Focusing on scholarship from the undergraduate level is consistent with a culture that values discovery and its application for the best care of patients. INSPIRE is the first research experience at MSSM to be part of the 4-year undergraduate medical curriculum. The program provides students with research skill building, 360-degree feedback, oversight, and mentorship. INSPIRE also offers training in clinical and translational research within the curriculum, thereby addressing the new Liaison Committee on Medical Education requirement that "the curriculum must introduce students to the basic principles of clinical and translational research, including how such research is conducted, evaluated, explained to patients, and applied to patient care."

INSPIRE has the potential to revamp the fourth-year medical student curriculum at MSSM, making it more meaningful and providing important additional competencies that will enrich our students' practice of medicine. This is reflected in the productivity of the participants, 3 of whom had peer-reviewed abstracts accepted for presentation and 1 of whom submitted a manuscript for publication. The program is in its infancy, but once the new curriculum for years 1 to 3 is in place, we plan to structure it as a required longitudinal track that begins early in students' education, perhaps as early as the summer following the first year, and one that fosters mastery of essential research skills. As such, it will transform our medical students' training. This, however, will require the development of medical student competencies in order to identify which students will be able to devote a year to INSPIRE and which will need that year for continued study. An initiative to define these competencies is currently underway at MSSM.

Including scholarly activities in undergraduate medical education offers students the means to experience expanded mentored research opportunities, to increase their awareness of the importance of critical thinking skills, to improve their mastery of the basic vocabulary and skills required to conduct research, and to help them understand why research is important for clinical care and how it is accomplished. This training may encourage students to pursue an academic career that includes a research

component, which may also address the reported decrease in R01s received by MDs.¹¹ The institution benefits as well by identifying and nurturing talented future physicians and physician-scholars. Scholarly programs such as INSPIRE may contribute to the continued recruitment of the best applicants and help to attract the best students to continue training at MSSM as residents, fellows, and faculty, enhancing departmental and institutional growth and enhancing Mount Sinai's academic and community missions.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Irby D. Four big ideas from the Carnegie Study on Medical Education. Available at: <http://www.youtube.com/watch?v=67bp-sk-mqc>. Accessed April 29 2009.
2. Association of American Medical Colleges. Promoting translational and clinical science: the critical role of medical schools and teaching hospitals. Report of the AAMC's Task Force II on Clinical Research. Available at: https://services.aamc.org/publications/index.cfm?fuseaction=product.displayform&_id=150&prv_id=176. Accessed April 29 2009.
3. Office of Student Research. Medical doctorate thesis requirements. Available at: <http://www.med.yale.edu/facres/thesis/index.shtml>. Accessed April 29 2009.
4. Schor NF, Troen P, Kanter SL, Levine AS. The Scholarly Project Initiative: introducing scholarship in medicine through a longitudinal, mentored curricular program. *Acad Med* 2005; 80: 824–831.
5. Stanford School of Medicine. Scholarly concentrations. Available at: http://med.stanford.edu/md/curriculum/scholarly_concentrations. Accessed April 29 2009.
6. Grochowski CO, Halperin EC, Buckley EG. A curricular model for the training of physician scientists: the evolution of the Duke University School of Medicine curriculum. *Acad Med* 2007; 82: 375–382.
7. UCSF School of Medicine. Pathways to discovery. Available at: <http://medschool.ucsf.edu/pathways>. Accessed April 29 2009.
8. Richards E, Borkan J, Gruppuso PA. Educating the next generation of leaders in medicine: the Scholarly Concentrations Program at the Warren Alpert Medical School of Brown University. *Med Health R I* 2007; 90: 275–276.
9. Penn Medicine. Scholarly study guidelines: general instructions. Available at: <http://www.med.upenn.edu/student/scholarlystudy>. Accessed April 29 2009.
10. Zier K, Friedman E, Smith L. Supportive programs increase medical students' research interest and productivity. *J Investig Med* 2006; 54: 201–207.
11. Rosenberg LE. Physician-scientists—endangered and essential. *Science* 1999; 283: 331–332.

Risk of Upper Aerodigestive Tract Cancer Associated with Smoking with and without Concurrent Alcohol Consumption

Alireza Ansary-Moghaddam, PhD,¹ Rachel R. Huxley, PhD,¹ Tai Hing Lam, PhD,² and Mark Woodward, PhD³

¹George Institute for International Health, New South Wales, Australia

²University of Hong Kong, Hong Kong, China

³Mount Sinai School of Medicine, New York, NY

ABSTRACT

Background: Smoking and alcohol are major causal factors for upper aerodigestive tract cancer, but reliable quantification of the combined impact of smoking and alcohol on this cancer and its major subtypes has not been performed.

Methods: A meta-analysis of studies that had published quantitative estimates of smoking and upper aerodigestive tract cancer by January 2007 was performed. Pooled estimates of relative risks were obtained. Publication bias was investigated through funnel plots and corrected if found to be present.

Results: Overall, 85 studies with information on 53,940 individuals with upper aerodigestive tract cancer were included. The pooled estimate for the association between smoking and the risk of this cancer was 3.47 (95% confidence interval, 3.06–3.92). The risk remained elevated for a decade after smoking cessation but declined thereafter. Individuals who both smoked and consumed alcohol had double the risk of upper aerodigestive tract cancer in comparison with those who only smoked: the relative risk was 6.93 (95% confidence interval,

4.99–9.62) for the former and 2.56 (95% confidence interval, 2.20–2.97) for the latter ($P < 0.001$).

Conclusions: Public health interventions that simultaneously discourage smoking and heavy drinking would have greater benefits than would be expected from those that target only one of these risk factors. *Mt Sinai J Med* 76:392–403, 2009. © 2009 Mount Sinai School of Medicine

Key Words: alcohol, cancer, meta-analysis, smoking, upper aerodigestive tract.

Upper aerodigestive tract cancer (UADTC; oropharyngeal, esophageal, and laryngeal) is an important global problem, with nearly 1 million new cases occurring annually, three-quarters of which are in developing countries.¹ Accordingly, in developing countries, UADTC ranks among the 5 most common cancers in both men and women.^{2,3} Additionally, the 5-year survival rates of some types of UADTC (eg, esophageal cancer) are among the lowest of all cancers at around 10%.^{2,3}

Cigarette smoking is a major causal factor for UADTC and its subtypes,⁴ explaining approximately 40% of all UADTC cancer deaths globally.^{5–7} Furthermore, alcohol consumption has been suggested to further exacerbate the effects of smoking on the risk of developing UADTC.⁶ However, quantification of the additional risk among smokers conferred by alcohol on specific subtypes of UADTC has been hindered by studies that have had too few events to enable any reliable estimation.⁶

Given the high and, in some regions, increasing prevalence of both smoking and alcohol consumption across large parts of the world, including developing countries,^{8–10} reliable quantification of the

Address Correspondence to:

Mark Woodward

Mount Sinai School of
Medicine, New York, NY
Email: mark.woodward@mssm.edu.

individual and combined effects of smoking and alcohol consumption on the risk of developing UADTC is warranted.

MATERIALS AND METHODS

Data Sources

Relevant studies were identified through EMBASE and MEDLINE with the MeSH terms (all exploded): "esophageal," "laryngeal," "pharyngeal," or "oral cavity" together with "neoplasm," "squamous cell," "adenocarcinoma," or "cancer" together with "cohort study" or "case-control study." The titles of our search results were reviewed, and clearly unrelated titles were excluded. Next, the abstracts of all possibly relevant studies were reviewed, and complete articles were obtained when the studies appeared relevant. References from identified studies were also scanned to identify any other relevant studies. We avoided duplicate studies by taking only the most recent publication with appropriate data.

Study Selection and Data Synthesis

Studies were included if they had published quantitative estimates and standard errors (or some other measure of variability) of the association between smoking and UADTC by January 2007. Studies were excluded if they provided only an estimate of effect, with no means by which to calculate the standard error, or if the estimates were not at least age-adjusted. Pooled estimates of relative risks (RRs) were obtained with either hazard ratios (HRs; for cohort studies) or odds ratios (ORs; for case-control studies) by means of a random effects analysis. Each study was weighted according to an estimate of its statistical size, which was defined as the inverse of the variance of log OR or HR.¹¹ Heterogeneity was estimated with the I^2 statistic and tested with the Q statistic.¹¹

Publication bias was investigated through funnel plots and formally assessed with Egger's test.^{11,12} Funnel plots are graphs of the exposure estimated from individual studies against a measure of the study size and are so called because the precision in the estimation of the underlying exposure increases as the size of the studies increases. If bias is absent, results from small studies will scatter relatively widely in comparison with those from larger studies, all around the same average. Corrections for publication bias were made with the trim-and-fill procedure, which essentially corrects the

funnel plot by imputing where the missing studies would most likely have fallen should they have been recorded.¹¹

In addition, studies (1 cohort and 7 case-control) that used similar categories for the level of daily cigarette consumption were pooled separately to examine any overall dose-response effect. Possible sources of heterogeneity were investigated by a comparison of the results of studies combined with respect to particular characteristics (eg, sex, study design, level of adjustment, and alcohol consumption) wherever possible. All analyses were performed with STATA, version 10.

Information on alcohol consumption was not reported in a consistent manner, but in several studies ($n = 10$), alcohol consumption was reported as "drinker" versus "never/almost never drinker." To examine the impact that alcohol consumption may have on the association between cigarette smoking and UADTC, the RR of UADTC among smokers who were reported to consume alcohol was compared with the risk estimate for smokers who were reported to be "never/almost never drinkers." Some studies included additional data on the quantity of alcohol and cigarettes consumed, enabling an exploration of the possible dose-response association between smoking, alcohol, and UADTC. From these, comparative data were obtained for heavy smokers who did not drink alcohol or who were reported to drink at a "moderate" level and for individuals that were reported to be both heavy smokers and drinkers.

RESULTS

A total of 15 cohort or nested case-control studies^{13–27} and 72 case-control studies,^{28–99} with information on a total of 53,940 individuals with UADTC (esophageal, 34,287; laryngeal, 4719; and oropharyngeal, 14,934; 41% were fatal), were included in the meta-analysis. The summary characteristics of included studies are shown in Supporting Tables 1 to 3 (in reverse chronological order; case-control studies are also grouped by cancer site). The majority of the studies were from Europe ($n = 30$), Asia ($n = 26$), and North America ($n = 18$), with the remaining 11 studies coming from South America, Africa, and Australia.

Smoking and the Risk of Fatal and Nonfatal Upper Aerodigestive Tract Cancer

The pooled estimate of the 120 identified reports for the association between current smoking and risk

of UADTC was 3.47 [95% confidence interval (CI), 3.06–3.92], with significant heterogeneity between the studies ($P < 0.001$). The summary estimate for the association between smoking and UADTC from studies that used nonsmokers as the referent group was significantly lower than that obtained from studies that used never smokers as the referent group: the RRs were 2.59 (95% CI, 2.09–3.21) and 3.74 (95% CI, 3.28–4.27), respectively (P for heterogeneity = 0.004). To avoid issues concerned with statistical analyses based on different reference groups, subsequent analyses were restricted to studies ($n = 94$) in which risk estimates were reported for current smokers compared with never smokers. After restriction of the analysis to these studies, there remained significant heterogeneity between the study estimates ($P < 0.001$) that was not explained by differences in study design: the summary estimates for case-control and cohort studies were an RR of 3.91 (95% CI, 3.36–4.53) and

an RR of 3.19 (95% CI, 2.42–4.20), respectively (P for heterogeneity = 0.20).

A substantial amount of the heterogeneity was explained by differences between the 3 subtypes of UADTC, with the strength of the association being significantly stronger for laryngeal cancer (RR, 9.07; 95% CI, 6.33–13.0; Figure 1) versus either oropharyngeal cancer (RR, 3.41; 95% CI, 2.74–4.24; $P < 0.001$; Figure 2) or esophageal cancer (RR, 3.07; 95% CI, 2.63–3.58; $P < 0.001$; Figure 3).

Geographical differences between the study populations were an additional source of heterogeneity between study estimates, with the magnitude of the association being significantly greater among populations from Europe and North America compared with Asia: the RRs were 4.05 (95% CI, 3.33–4.92) and 2.64 (95% CI, 2.19–3.17), respectively (P for heterogeneity = 0.02).

Among the 12 studies that reported sex-specific effects, there was no evidence to suggest that the

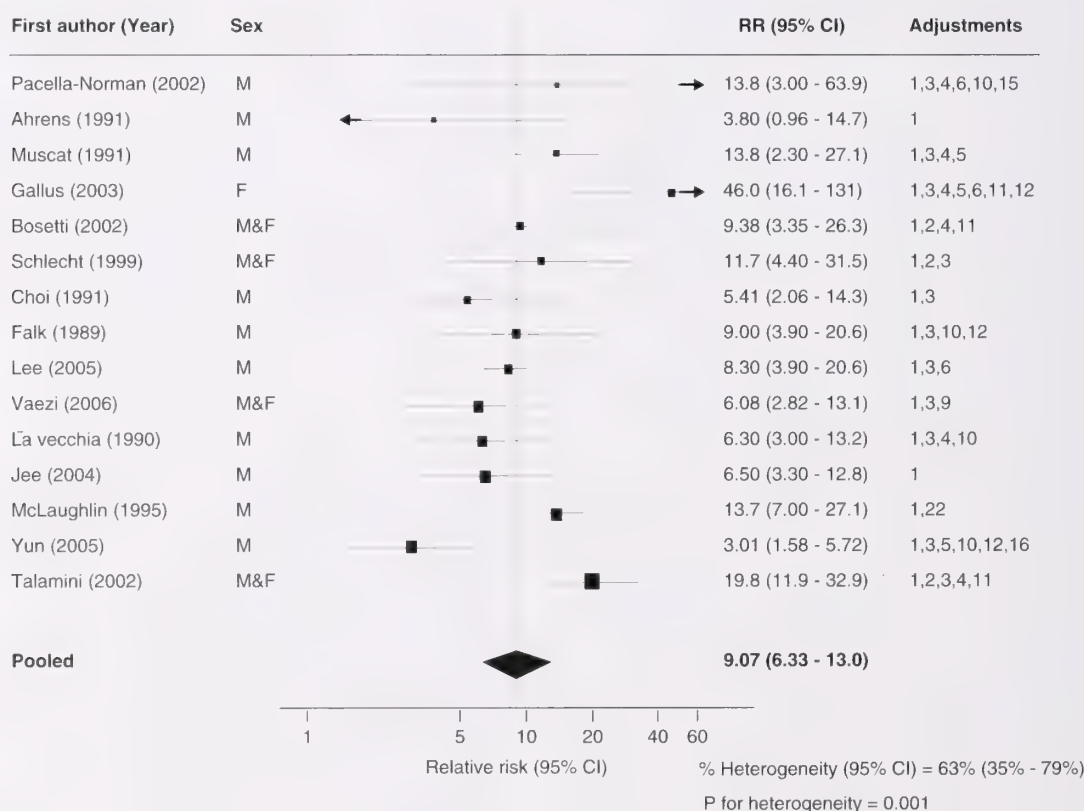


Fig 1. Smoking and laryngeal cancer in a meta-analysis. Black squares indicate point estimates (with the area proportional to statistical information based on the inverse of the variance of the RR provided by each study), and horizontal lines indicate 95% CIs for the observed effect in each study. A diamond indicates a pooled estimate and 95% CI for meta-analysis. The levels of adjustment are as follows: (1) age, (2) sex, (3) alcohol, (4) education, (5) body mass index, (6) other types of tobacco, (7) tea, (8) coffee, (9) other related disease, (10) socioeconomic status, (11) study center, (12) diet, (13) marital status, (14) race, (15) place of birth, (16) physical activity, (17) religion, (18) family history, (19) blood pressure, (20) triglyceride, (21) blood cholesterol, and (22) observation period. **Abbreviations:** CI, confidence interval; F, female; M, male; RR, relative risk.

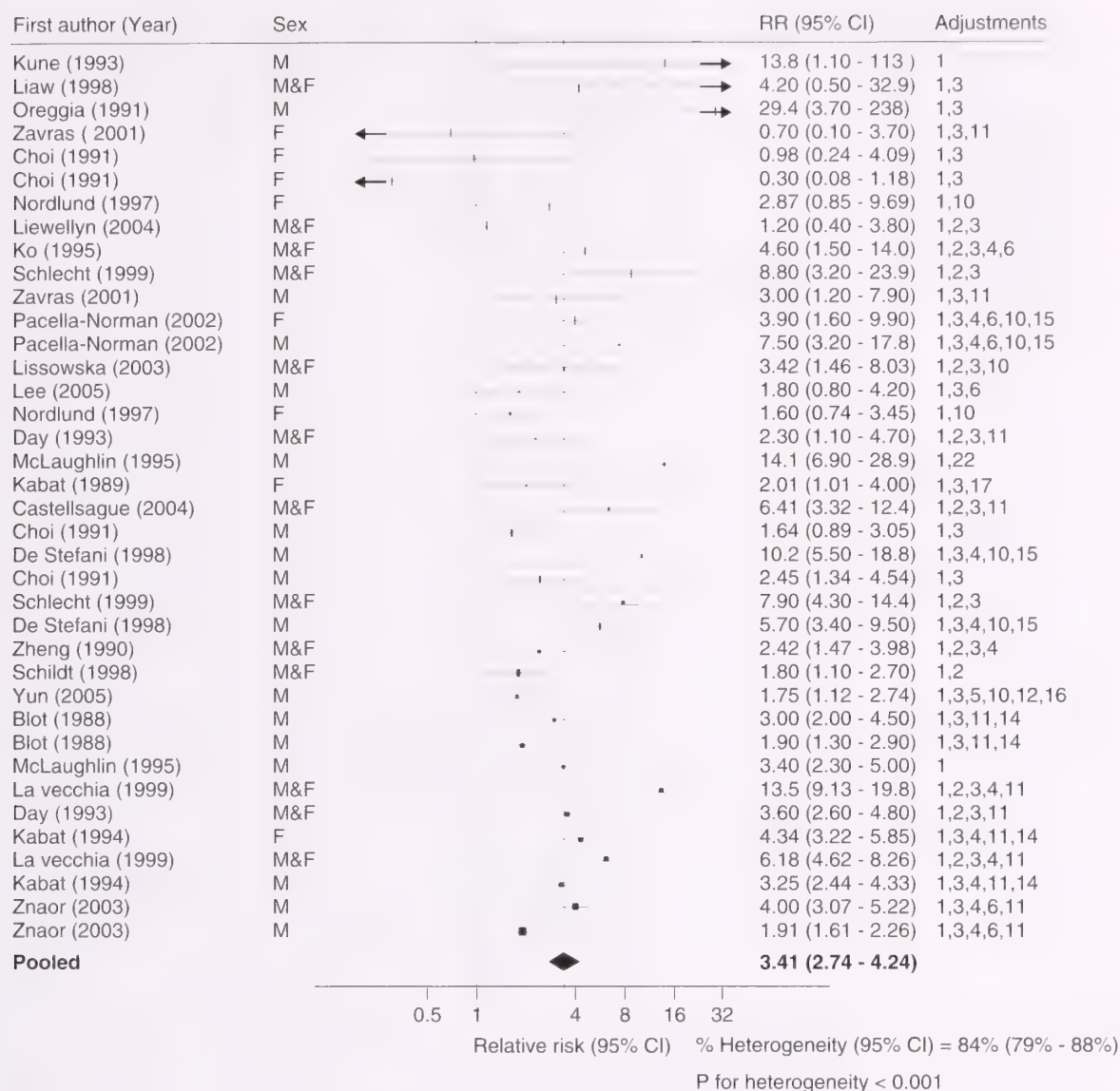


Fig 2. Smoking and oropharyngeal cancer in a meta-analysis (for definitions of the symbols, lines, and adjustment levels, see Figure 1). **Abbreviations:** CI, confidence interval; F, female; M, male; RR, relative risk.

association between smoking and UADTC differed between men and women: the RR was 2.81 (95% CI, 2.13–3.70) for men and 2.62 (95% CI, 1.78–3.80) for women (P for heterogeneity = 0.77).

Impact of Publication Bias on the Strength of the Association

There was evidence of publication bias for the association between current smoking (versus never smoking) and UADTC ($P = 0.002$). Subsequent trim-and-fill analysis attenuated the association, such that the RR was reduced from 3.74 to 2.89 (95% CI, 2.50–3.34). Correction for the presence of publication bias also altered the risk estimates for the individual subgroups: the RR was 2.34 (95% CI, 1.75–3.13) in

cohort studies and 3.14 (2.67–3.69) in case-control studies (P for heterogeneity = 0.08). The difference between regions was also reduced and became nonsignificant: the RR was 2.39 (95% CI, 1.98–2.89) in Asian countries and 3.10 (95% CI, 2.50–3.85) in non-Asian countries (P for heterogeneity = 0.076).

Within subtypes of UADTC, publication bias was observed for studies reporting on the association between smoking and esophageal cancer ($P = 0.016$) but not for cancers of the larynx ($P = 0.92$) or oropharynx ($P = 0.54$). After correction for the presence of publication bias, the RR for cancer of the esophagus associated with smoking was attenuated to 2.52 (95% CI, 2.14–2.95).

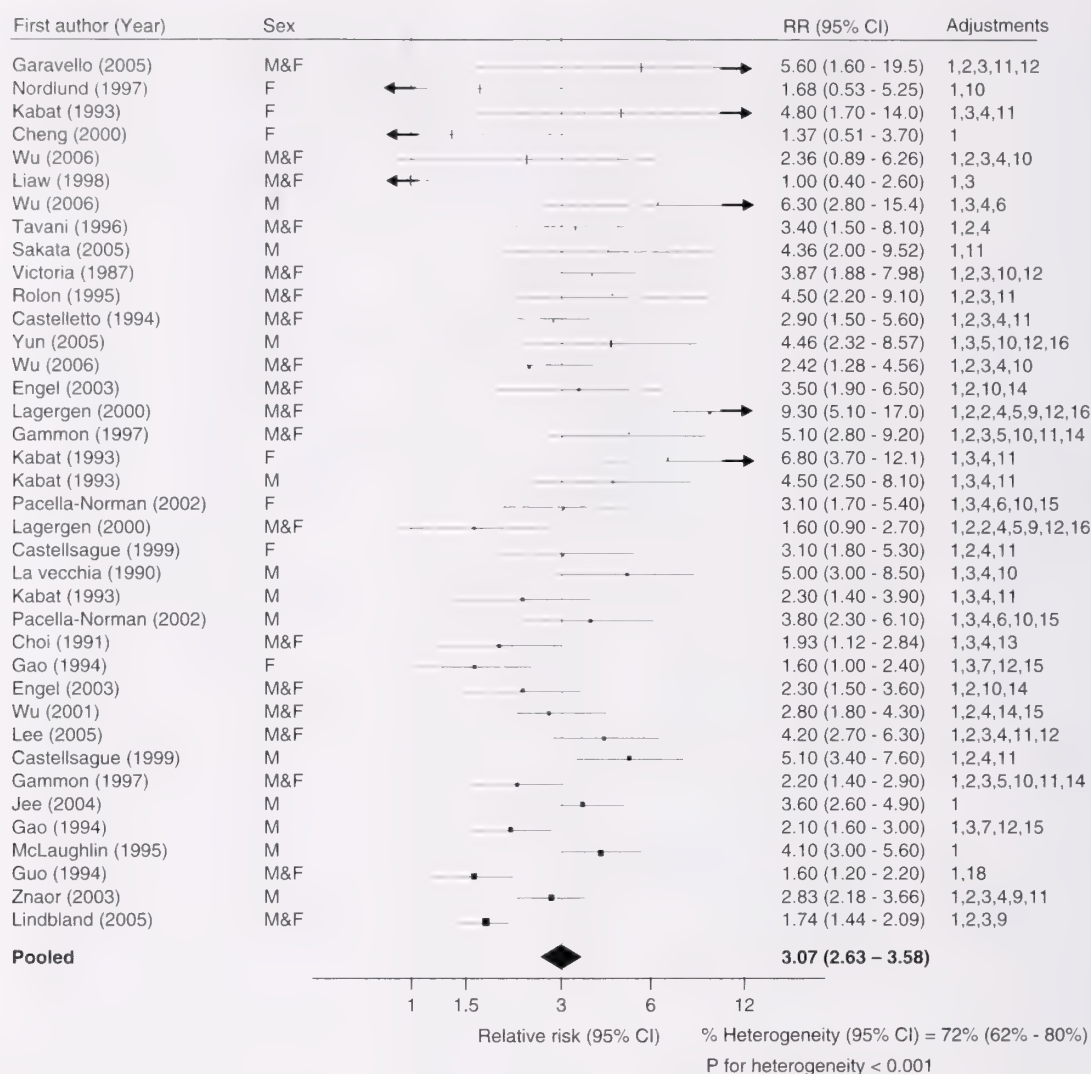


Fig 3. Smoking and esophageal cancer in a meta-analysis (for definitions of the symbols, lines, and adjustment levels, see Figure 1). **Abbreviations:** CI, confidence interval; F, female; M, male; RR, relative risk.

Dose-Response Relationship Between Cigarette Smoking and Upper Aerodigestive Tract Cancer

A total of 15 studies had reported 3 mutually exclusive categories for daily cigarette consumption: never smokers, <20 cigarettes per day, and ≥ 20 cigarettes per day. A strong dose-response association between the number of cigarettes smoked per day and the risk of UADTC was observed, such that individuals who smoked 20 or more cigarettes a day had 4 times the risk of UADTC versus never smokers (RR, 4.42; 95% CI, 3.07–6.36; Table 1), which remained unchanged after correction for publication bias. A dose-response relationship was consistently shown for each of the 3 subtypes of UADTC both before and after adjustment for publication bias (Table 1).

Smoking Cessation and Risk of Upper Aerodigestive Tract Cancer in Former Smokers

Data were available on the relationship between quitting smoking and the risk of UADTC from 80 reports, which analyzed 2,061,322 individuals and 18,234 events. Compared with never smokers, former smokers had double the risk of UADTC, which was approximately half that of current smokers (Table 1). The risk of UADTC remained elevated in the first decade after smoking cessation but declined thereafter; 10 years and more after smoking cessation, the risk of UADTC among former smokers compared with never smokers was nonsignificantly higher: the RR was 1.23 (95% CI, 0.82–1.84). There was again evidence of publication bias ($P = 0.02$), such that after correction, the RR of UADTC among former

Table 1. Association Between Cigarette Smoking and UADTC in a Meta-Analysis.

Smoking Habit Cancer Type	Number		RR (95% CI)	
	Reports	Individual Events	Multiple Adjusted	Corrected for Publication Bias
All UADTC				
Never smoking			1	1
Former smokers	80	18,234	1.92 (1.62–2.29)	1.29 (1.09–1.52)
Current smokers	94	22,884	3.74 (3.28–4.27)	2.89 (2.50–3.34)
Dose response				
All UADTC				
Never smoking			1	1
<20 CPD	15	2104	2.49 (1.76–3.51)	2.49 (1.76–3.51)
≥ 20 CPD			4.42 (3.07–6.36)	4.42 (3.07–6.36)
Esophagus				
Never smoking			1	1
<20 CPD	7	685	2.44 (1.87–3.20)	2.36 (1.66–3.37)
≥ 20 CPD			3.93 (2.68–5.75)	2.97 (2.01–4.40)
Oropharynx				
Never smoking			1	1
<20 CPD	6	1097	1.61 (0.90–2.88)	1.61 (0.90–2.88)
≥ 20 CPD			3.12 (1.78–5.47)	3.12 (1.78–5.47)
Larynx				
Never smoking			1	1
<20 CPD	2	322	7.66 (4.09–14.3)	7.66 (4.09–14.3)
≥ 20 CPD			19.0 (8.86–40.4)	11.3 (3.02–42.3)
All UADTC: years since quitting smoking				
Never smoking			1	1
<10	11	878	3.39 (2.22–5.18)	2.74 (1.82–4.13)
≥ 10			1.23 (0.82–1.84)	1.04 (0.69–1.58)

NOTE: The referent group is the never smoking group.

Abbreviations: CI, confidence interval; CPD, cigarettes per day; RR, relative risk; UADTC, upper aerodigestive tract cancer.

smokers compared with never smokers was 1.29 (95% CI, 1.09–1.52), and 10 years after smoking cessation, the RR approached that of a never smoker: 1.04 (95% CI, 0.69–1.58).

Combined Effects of Alcohol and Smoking on the Risk of Upper Aerodigestive Tract Cancer

A total of 21 reports estimated the effect of exposure to both alcohol and smoking on the risk of UADTC (Figure 4). The RR of UADTC among individuals who both smoked and consumed alcohol was considerably greater than that for individuals who only smoked: the RRs were 6.93 (95% CI, 4.99–9.62) and 2.56 (95% CI, 2.20–2.97), respectively ($P < 0.001$). For all UADTC subtypes, alcohol consumption exacerbated the association between smoking and cancer (Figure 4). After correction for publication bias, the RR for UADTC was attenuated to 5.26 (95% CI, 3.69–7.50) in subjects who both smoked and consumed alcohol in comparison with current smokers who were nondrinkers.

A total of 24 studies (29 reports) examined the dose-response relationship between smoking and alcohol consumption with the risk of UADTC. The RR of UADTC among individuals who had the highest reported consumption of both cigarettes and alcohol was close to 5 times the risk among heavy smokers who did not drink or drank at a moderate level: the RRs were 29.3 (95% CI, 19.5–43.9) and 5.76 (95% CI, 4.10–8.08), respectively ($P < 0.001$). The joint effects of smoking and alcohol varied by the cancer subtype: the RR was 10.0 (95% CI, 4.08–24.5) for the esophagus, 35.6 (95% CI, 23.6–53.7) for the oropharynx, and 60.5 (95% CI, 19.4–188) for the larynx. After correction for the presence of publication bias, the risk of developing UADTC for individuals with a high consumption of both cigarettes and alcohol in comparison with individuals who were heavy smokers but nondrinkers was reduced. In comparison with heavy smokers and nondrinkers, the RRs of UADTC and its individual subtypes (esophageal, oropharyngeal, and laryngeal) among heavy smokers and drinkers were 15.6 (95% CI, 10.2–23.9), 5.93 (95% CI,

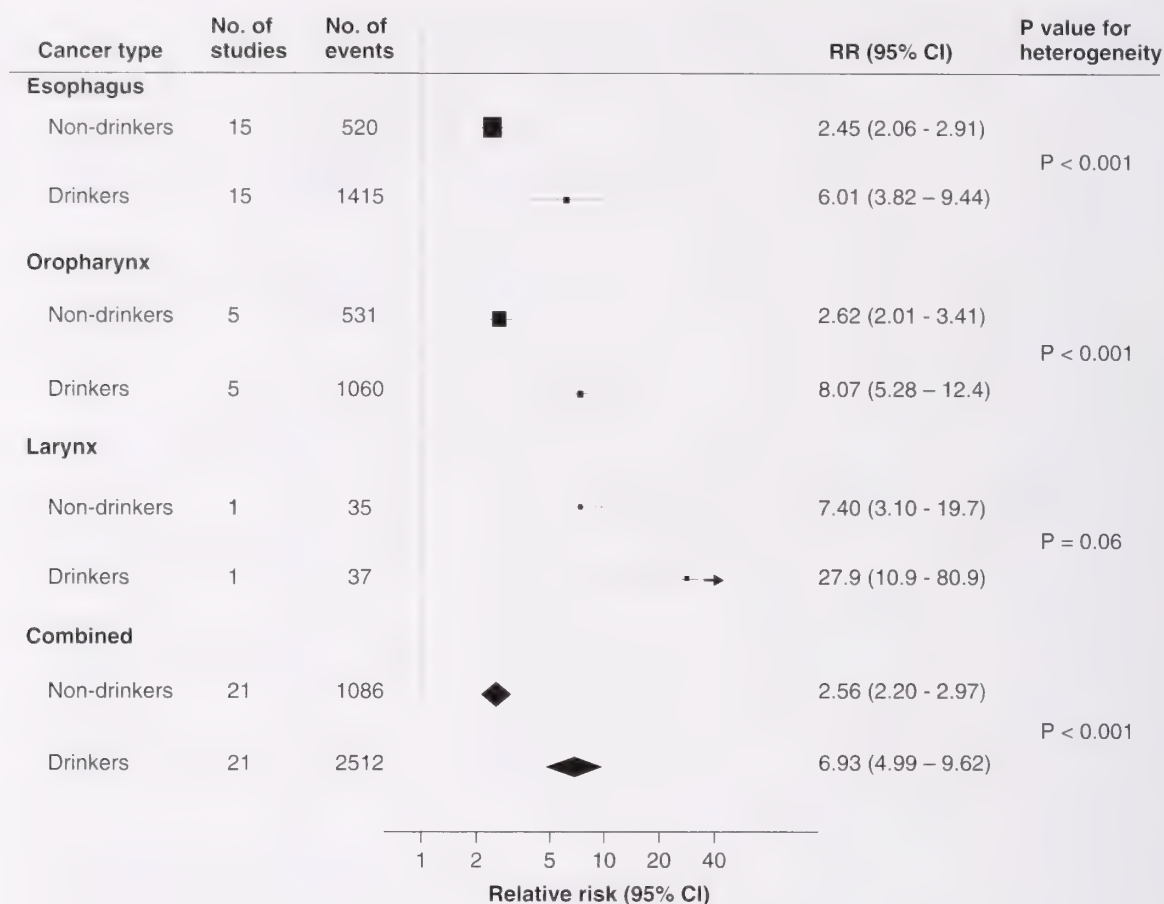


Fig 4. Association between smoking and upper aerodigestive tract cancer by alcohol status in a meta-analysis (for definitions of the symbols and lines, see Figure 1). **Abbreviations:** CI, confidence interval; RR, relative risk.

2.46–14.3), 29.4 (95% CI, 18.9–45.9), and 17.5 (95% CI, 5.37–56.7), respectively,

DISCUSSION

Findings from this meta-analysis of more than 50,000 UADTC events from 85 studies confirm the strong dose-response association between cigarette smoking and UADTC and its 3 major subtypes for both men and women. The risks were particularly strong for laryngeal cancer compared with cancers of the esophagus and oropharynx. In agreement with previous reports,^{6,100} these findings clearly indicate that the risks associated with smoking are amplified by alcohol, with the greatest hazards for UADTC occurring among individuals who were in the highest categories for both smoking and alcohol consumption. It is of particular importance to public health that quitting smoking substantially lessens the chance of developing UADTC, especially among former smokers who have ceased smoking for 10 or more years. Although this is not a novel finding, the

magnitude of the effects of smoking and drinking has never previously been quantified so reliably, including pooling of all available data and corrections for bias caused by underreporting.

The mechanism that may underpin the interaction between smoking and alcohol consumption in the etiology of UADTC remains unclear. Ethanol itself is carcinogenic agent¹⁰¹; furthermore, its primary metabolite, acetaldehyde (which is also present in high concentrations in cigarette smoke), has been shown in animal studies to be tumor-promoting.¹⁰² Hence, increased exposure to acetaldehyde from the simultaneous smoking of cigarettes and consumption of alcoholic beverages may explain the greatly increased risk of UADTC observed among heavy smokers and drinkers.

There was some suggestion that the association between smoking and UADTC was weaker among Asian populations in comparison with the associations from Western studies, although after correction for publication bias, there was no longer any regional difference. This may be due to differences in the compositions of the cigarettes sold in Asia and the

West as well as differences in smoking habits, such as frequency and depth of inhalation. However, given the substantial differences in the maturity of the smoking epidemic between the regions, it would not have been unexpected to have observed smaller estimates of effect among Asian populations compared with Western populations. Any difference in risk is, however, likely to be an artifact rather than a reflection of any substantive genetic or physiological differences that might confer some protective effect against smoking in Asians, as has been previously speculated.¹⁰³ Instead, it more likely reflects the fact that there is usually a time lag of several decades between commencement of smoking and disease onset; hence, in countries such as China, where smoking did not reach its peak popularity until the mid-1990s,¹⁰⁴ the full impact of smoking has yet to be realized.

There are several potential limitations that warrant discussion. The majority of identified studies were hospital-based case-control studies and consequently may have been prone to recall and selection bias. In addition, different exposure definitions/groups made it difficult to include all studies in subgroup analyses. Some studies did not exclude former smokers from the reference group (never smokers). The individual studies included in the meta-analysis also used different methods to verify UADTC, and this could have had some unpredictable effect on the results. However, misdiagnosis of these cancers is unlikely.

To our knowledge, this is the most comprehensive review of observational studies on smoking as the most important risk factor for UADTC both independently and in conjunction with alcohol consumption. The great strength of this analysis is the large number of events, which facilitated our ability to assess associations by subgroups. Findings from this study have clear public health implications. Tackling smoking and alcohol use in the populous Asia-Pacific region could prevent millions of deaths from UADTC as well as other smoking-related and alcohol-related outcomes over the next few decades. Therefore, a population-wide approach that reduced the prevalence of smoking and alcohol together would have a significant impact on reducing the burden of UADTC worldwide and, in particular, in economically developing countries of the Asia-Pacific region.

LATE BREAKING RESEARCH

Since we concluded our systematic review, several new studies of relevance have been published.^{105–116} In addition, our findings receive confirmatory support from a recent meta-analysis from the International Head and Neck Cancer Epidemiology Consortium.¹¹⁷ In that study, which involved 11,000 cases of head and neck cancer and 16,000 controls, the risk of head and neck cancer among ever tobacco users and alcohol consumers was also approximately 6 times greater than that among never smokers and nondrinkers, with evidence of a dose-response relation equivalent to that which we report in the current review.

Acknowledgements:

This project has received grants from the National Health and Medical Research Council of Australia. Alireza Ansary-Moghaddam received a scholarship from Zahedan University of Medical Sciences (Iran).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Ferlay J, Bray F, Pisani P, Parkin DM. *GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. IARC Cancer Base No 5. Version 2.0*. Lyon, France: IARC Press; 2004.
2. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of eighteen major cancers in 1985. *Int J Cancer* 1993; 54: 594–606.
3. Schottenfeld D, Fraumeni F, eds. *Cancer Epidemiology and Prevention*. 3rd ed. New York, NY: Oxford University Press; 2006.
4. *Tobacco Smoke and Involuntary Smoking*. Lyon, France: International Agency for Research on Cancer; 2004; *IARC Monographs on the Evaluation of Carcinogenic Risk to Humans*; vol 83.
5. Ezzati M, Henley SJ, Lopez AD, et al. Role of smoking in global and regional cancer epidemiology: current patterns and data needs. *Int J Cancer* 2005; 116: 963–971.

6. Zeka A, Gore R, Kriebel D. Effects of alcohol and tobacco on aero digestive cancer risks: a meta-regression analysis. *Cancer Causes Control* 2003; 14: 897–906.
7. Gandini S, Botteri E, Iodice S, et al. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer* 2008; 122: 155–164.
8. World Health Organization. WHO global status report on alcohol 2004. Available at: http://search.who.int/search?ie=utf8&site=default_collection&client=who&proxystylesheet=who&output=xml_no_dtd&oe=utf8&=alcohol. Accessed May 2009.
9. Jha P, Paccaud F, Nguyen S. Strategic priorities in tobacco control for governments and international agencies. Available at: <http://www1.worldbank.org/tobacco/tcdc/449to464.pdf>. Accessed May 2009.
10. Martiniuk AL, Lee CM, Lam TH, et al. The fraction of ischaemic heart disease and stroke attributable to smoking in the WHO Western Pacific and South-East Asian regions. *Tob Control* 2006; 15: 181–188.
11. Woodward M. *Epidemiology: Study Design and Data Analysis*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC 2005.
12. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629–634.
13. Ishikawa A, Kuriyama S, Tsubono Y, et al. Smoking, alcohol drinking, green tea consumption and the risk of esophageal cancer in Japanese men. *J Epidemiol* 2006; 16: 185–192.
14. Yun YH, Jung KW, Bae JM, et al. Cigarette smoking and cancer incidence risk in adult men: National Health Insurance Corporation study. *Cancer Detect Prev* 2005; 29: 15–24.
15. Tran GD, Sun XD, Abnet CC, et al. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int J Cancer* 2005; 113: 456–463.
16. Lindblad M, Rodriguez LA, Lagergren J. Body mass, tobacco and alcohol and risk of esophageal, gastric cardia, and gastric non-cardia adenocarcinoma among men and women in a nested case-control study. *Cancer Causes Control* 2005; 16: 285–294.
17. Sakata K, Hoshiyama Y, Morioka S, et al., for the JACC Study Group. Smoking, alcohol drinking and esophageal cancer: findings from the JACC study. *J Epidemiol* 2005; 15: S212–S219.
18. Jee SH, Samet JM, Ohrr H, et al. Smoking and cancer risk in Korean men and women. *Cancer Causes Control* 2004; 15: 341–348.
19. Nilsson S, Carstensen JM, Pershagen G. Mortality among male and female smokers in Sweden: a 33 year follow-up. *J Epidemiol Community Health* 2001; 55: 825–830.
20. Liaw KM, Chen CJ. Mortality attributable to cigarette smoking in Taiwan: a 12-year follow-up study. *Tob Control* 1998; 7: 141–148.
21. Lam TH, He Y, Li LS, et al. Mortality attributable to cigarette smoking in China. *JAMA* 1997; 278: 1505–1508.
22. Nordlund LA, Carstensen JM, Pershagen G. Cancer incidence in female smokers: a 26-year follow-up. *Int J Cancer* 1997; 73: 625–628.
23. Raitiola HS, Pukander JS. Etiological factors of laryngeal cancer. *Acta Otolaryngol Suppl* 1997; 529: 215–217.
24. Engeland A, Andersen A, Haldorsen T, et al. Smoking habits and risk of cancers other than lung cancer: 28 years' follow-up of 26,000 Norwegian men and women. *Cancer Causes Control* 1996; 7: 497–506.
25. Chyou PH, Nomura AM, Stemmermann GN. Diet, alcohol, smoking and cancer of the upper aerodigestive tract: a prospective study among Hawaii Japanese men. *Int J Cancer* 1995; 60: 616–621.
26. McLaughlin JK, Hrubec Z, Blot WJ, et al. Smoking and cancer mortality among U.S. veterans: a 26-year follow-up. *Int J Cancer* 1995; 60: 190–193.
27. Guo W, Blot WJ, Li JY, et al. A nested case-control study of oesophageal and stomach cancers in the Linxian nutrition intervention trial. *Int J Epidemiol* 1994; 23: 444–450.
28. Wu IC, Lu CY, Kuo FC, et al. Interaction between cigarette, alcohol and betel nut use on esophageal cancer risk in Taiwan. *Eur J Clin Invest* 2006; 36: 236–241.
29. Wu M, Zhao JK, Hu XS, et al. Association of smoking, alcohol drinking and dietary factors with esophageal cancer in high- and low-risk areas of Jiangsu province, China. *World J Gastroenterol* 2006; 12: 1686–1693.
30. Jiang JM, Zeng XJ, Chen JS, et al. Smoking and mortality from esophageal cancer in China: a large case-control study of 19,734 male esophageal cancer deaths and 104,846 living spouse controls. *Int J Cancer* 2006; 119: 1427–1432.
31. Yang CX, Wang HY, Wang ZM, et al. Risk factors for esophageal cancer: a case-control study in South-Western China. *Asian Pac J Cancer Prev* 2005; 6: 48–53.
32. Lee CH, Lee JM, Wu DC, et al. Independent and combined effects of alcohol intake, tobacco smoking and betel quid chewing on the risk of esophageal cancer in Taiwan. *Int J Cancer* 2005; 113: 475–482.
33. Garavello W, Negri E, Talamini R, et al. Family history of cancer, its combination with smoking and drinking, and risk of squamous cell carcinoma of the esophagus. *Cancer Epidemiol Biomarkers Prev* 2005; 14: 1390–1393.
34. Engel LS, Chow WH, Vaughan TL, et al. Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst* 2003; 95: 1404–1413.
35. Znaor A, Brennan P, Gajalakshmi V, et al. Independent and combined effects of tobacco smoking, chewing and alcohol drinking on the risk of oral, pharyngeal and esophageal cancers in Indian men. *Int J Cancer* 2003; 105: 681–686.
36. Pacella-Norman R, Urban MI, Sitas F, et al. Risk factors for oesophageal, lung, oral and laryngeal cancers in black South Africans. *Br J Cancer* 2002; 86: 1751–1756.
37. Wu AH, Wan P, Bernstein L. A multiethnic population-based study of smoking, alcohol and body size and risk of adenocarcinomas of the stomach and esophagus (United States). *Cancer Causes Control* 2001; 12: 721–732.
38. Lam TH, Ho SY, Hedley AJ, et al. Mortality and smoking in Hong Kong: case-control study of all adult deaths in 1998. *BMJ* 2001; 323: 1–6.
39. Lagergren J, Bergstrom R, Lindgren A, et al. The role of tobacco, snuff and alcohol use in the aetiology

- of cancer of the oesophagus and gastric cardia. *Int J Cancer* 2000; 85: 340-346.
40. Zamboni P, Talamini R, La Vecchia C, et al. Smoking, type of alcoholic beverage and squamous-cell oesophageal cancer in northern Italy. *Int J Cancer* 2000; 86: 144-149.
 41. Cheng KK, Sharp L, McKinney PA, et al. A case-control study of oesophageal adenocarcinoma in women: a preventable disease. *Br J Cancer* 2000; 83: 127-132.
 42. Castellsague X, Munoz N, De Stefani E, et al. Independent and joint effects of tobacco smoking and alcohol drinking on the risk of esophageal cancer in men and women. *Int J Cancer* 1999; 82: 657-664.
 43. Gammon MD, Schoenberg JB, Ahsan H, et al. Tobacco, alcohol, and socioeconomic status and adenocarcinomas of the esophagus and gastric cardia. *J Natl Cancer Inst* 1999; 89: 1277-1284.
 44. Tavani A, Negri E, Franceschi S, et al. Tobacco and other risk factors for oesophageal cancer in alcohol non-drinkers. *Eur J Cancer Prev* 1996; 5: 313-318.
 45. Rolon PA, Castellsague X, Benz M, et al. Hot and cold mate drinking and esophageal cancer in Paraguay. *Cancer Epidemiol Biomarkers Prev* 1995; 4: 595-605.
 46. Gao YT, McLaughlin JK, Blot WJ, et al. Risk factors for esophageal cancer in Shanghai, China. I. Role of cigarette smoking and alcohol drinking. *Int J Cancer* 1994; 58: 192-196.
 47. Brown LM, Silverman DT, Pottern LM, et al. Adenocarcinoma of the esophagus and esophagogastric junction in white men in the United States: alcohol, tobacco, and socioeconomic factors. *Cancer Causes Control* 1994; 5: 333-340.
 48. Castelletto R, Castellsague X, Munoz N, et al. Alcohol, tobacco, diet, mate drinking, and esophageal cancer in Argentina. *Cancer Epidemiol Biomarkers Prev* 1994; 3: 557-564.
 49. Tavani A, Negri E, Franceschi S, et al. Risk factors for esophageal cancer in women in northern Italy. *Cancer* 1993; 72: 2531-2536.
 50. Kabat GC, Ng SK, Wynder EL. Tobacco, alcohol intake, and diet in relation to adenocarcinoma of the esophagus and gastric cardia. *Cancer Causes Control* 1993; 4: 123-132.
 51. Choi SY, Kahyo H. Effect of cigarette smoking and alcohol consumption in the etiology of cancers of the digestive tract. *Int J Cancer* 1991; 49: 381-386.
 52. Sankaranarayanan R, Duffy SW, Padmakumary G, et al. Risk factors for cancer of the oesophagus in Kerala, India. *Int J Cancer* 1991; 49: 485-489.
 53. La Vecchia C, Bidoli E, Barra S, et al. Type of cigarettes and cancers of the upper digestive and respiratory tract. *Cancer Causes Control* 1990; 1: 69-74.
 54. Li JY, Ershow AG, Chen ZJ, et al. A case-control study of cancer of the esophagus and gastric cardia in Linxian. *Int J Cancer* 1989; 43: 755-761.
 55. Victora CG, Muñoz N, Day NE, et al. Hot beverages and oesophageal cancer in southern Brazil: a case-control study. *Int J Cancer* 1987; 39: 710-716.
 56. Vaezi MF, Qadeer MA, Lopez R, Colabianchi N. Laryngeal cancer and gastroesophageal reflux disease: a case-control study. *Am J Med* 2006; 119: 768-776.
 57. Lee KW, Kuo WR, Tsai SM, et al. Different impact from betel quid, alcohol and cigarette: risk factors for pharyngeal and laryngeal cancer. *Int J Cancer* 2005; 117: 831-836.
 58. Gallus S, Bosetti C, Franceschi S, et al. Laryngeal cancer in women: tobacco, alcohol, nutritional, and hormonal factors. *Cancer Epidemiol Biomarkers Prev* 2003; 12: 514-517.
 59. Talamini R, Bosetti C, La Vecchia C, et al. Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study. *Cancer Causes Control* 2002; 13: 957-964.
 60. Bosetti C, Gallus S, Franceschi S, et al. Cancer of the larynx in non-smoking alcohol drinkers and in non-drinking tobacco smokers. *Br J Cancer* 2002; 87: 516-518.
 61. Schlecht NF, Franco EL, Pintos J, et al. Effect of smoking cessation and tobacco type on the risk of cancers of the upper aero-digestive tract in Brazil. *Epidemiology* 1999; 10: 412-418.
 62. Rao DN, Desai PB, Ganesh B. Alcohol as an additional risk factor in laryngopharyngeal cancer in Mumbai—a case-control study. *Cancer Detect Prev* 1999; 23: 37-44.
 63. Dosemeci M, Gokmen I, Unsal M, et al. Tobacco, alcohol use, and risks of laryngeal and lung cancer by subsite and histologic type in Turkey. *Cancer Causes Control* 1997; 8: 729-737.
 64. Zheng W, Blot WJ, Shu XO, et al. Diet and other risk factors for laryngeal cancer in Shanghai, China. *Am J Epidemiol* 1992; 136: 178-191.
 65. Ahrens W, Jöckel KH, Patzak W, et al. Alcohol, smoking, and occupational factors in cancer of the larynx: a case-control study. *Am J Ind Med* 1991; 20: 477-493.
 66. Muscat JE, Wynder EL. Tobacco, alcohol, asbestos and occupational risk factors for laryngeal cancer. *Cancer* 1992; 69: 2244-2251.
 67. Franceschi S, Talamini R, Barra S, et al. Smoking and drinking in relation to cancers of the oral cavity, pharynx, larynx, and esophagus in northern Italy. *Cancer Res* 1990; 50: 6502-6507.
 68. Sankaranarayanan R, Duffy SW, Nair MK, et al. Tobacco and alcohol as risk factors in cancer of the larynx in Kerala, India. *Int J Cancer* 1990; 45: 879-882.
 69. Falk RT, Pickle LW, Brown LM, et al. Effect of smoking and alcohol consumption on laryngeal cancer risk in coastal Texas. *Cancer Res* 1989; 49: 4024-4029.
 70. Tuyns AJ, Estève J, Raymond L, et al. Cancer of the larynx/hypopharynx, tobacco and alcohol: IARC international case-control study in Turin and Varese (Italy), Zaragoza and Navarra (Spain), Geneva (Switzerland) and Calvados (France). *Int J Cancer* 1988; 41: 483-491.
 71. Brownson RC, Chang JC. Exposure to alcohol and tobacco and the risk of laryngeal cancer. *Arch Environ Health* 1987; 42: 192-196.
 72. Rodriguez T, Altieri A, Chatenoud L, et al. Risk factors for oral and pharyngeal cancer in young adults. *Oral Oncol* 2004; 40: 207-213.
 73. Liewellyn CD, Johnson NW, Warnakulasuriya KA. Risk factors for oral cancer in newly diagnosed patients aged 45 years and younger: a case-control study in Southern England. *J Oral Pathol Med* 2004; 33: 525-532.
 74. Castellsague X, Quintana MJ, Martinez MC, et al. The role of type of tobacco and type of alcoholic

- beverage in oral carcinogenesis. *Int J Cancer* 2004; 108: 741–749.
75. Lissowska J, Pilarska A, Pilarski P, et al. Smoking, alcohol, diet, dentition and sexual practices in the epidemiology of oral cancer in Poland. *Eur J Cancer Prev* 2003; 12: 25–33.
76. Garrote LF, Herrero R, Reyes RM, et al. Risk factors for cancer of the oral cavity and oro-pharynx in Cuba. *Br J Cancer* 2001; 85: 46–54.
77. Zavras AI, Douglass CW, Joshipura K, et al. Smoking and alcohol in the etiology of oral cancer: gender-specific risk profiles in the south of Greece. *Oral Oncol* 2001; 37: 28–35.
78. Hayes RB, Bravo-Otero E, Kleinman DV, et al. Tobacco and alcohol use and oral cancer in Puerto Rico. *Cancer Causes Control* 1999; 10: 27–33.
79. La Vecchia C, Franceschi S, Bosetti C, et al. Time since stopping smoking and the risk of oral and pharyngeal cancers. *J Natl Cancer Inst* 1999; 91: 726–728.
80. De Stefani E, Boffetta P, Oreggia F, et al. Smoking patterns and cancer of the oral cavity and pharynx: a case-control study in Uruguay. *Oral Oncol* 1998; 34: 340–346.
81. Schildt EB, Eriksson M, Hardell L, et al. Oral snuff, smoking habits and alcohol consumption in relation to oral cancer in a Swedish case-control study. *Int J Cancer* 1998; 77: 341–346.
82. Ko YC, Huang YL, Lee CH, et al. Betel quid chewing, cigarette smoking and alcohol consumption related to oral cancer in Taiwan. *J Oral Pathol Med* 1995; 24: 450–453.
83. Bundgaard T, Wildt J, Frydenberg M, et al. Case-control study of squamous cell cancer of the oral cavity in Denmark. *Cancer Causes Control* 1995; 6: 57–67.
84. Kabat GC, Chang CJ, Wynder EL. The role of tobacco, alcohol use, and body mass index in oral and pharyngeal cancer. *Int J Epidemiol* 1994; 23: 1137–1144.
85. Rao DN, Ganesh B, Rao RS, et al. Risk assessment of tobacco, alcohol and diet in oral cancer—a case-control study. *Int J Cancer* 1994; 58: 469–473.
86. Kune GA, Kune S, Field B, et al. Oral and pharyngeal cancer, diet, smoking, alcohol, and serum vitamin A and beta-carotene levels: a case-control study in men. *Nutr Cancer* 1993; 20: 61–70.
87. Day GL, Blot WJ, Austin DF, et al. Racial differences in risk of oral and pharyngeal cancer: alcohol, tobacco, and other determinants. *J Natl Cancer Inst* 1993; 85: 465–473.
88. Oreggia F, De Stefani E, Correa P, et al. Risk factors for cancer of the tongue in Uruguay. *Cancer* 1991; 67: 180–183.
89. Nandakumar A, Thimmasetty KT, Sreeramareddy NM, et al. A population-based case-control investigation on cancers of the oral cavity in Bangalore, India. *Br J Cancer* 1990; 62: 847–851.
90. Zheng TZ, Boyle P, Hu HF, et al. Tobacco smoking, alcohol consumption, and risk of oral cancer: a case-control study in Beijing, People's Republic of China. *Cancer Causes Control* 1990; 1: 173–179.
91. Kabat GC, Hebert JR, Wynder EL. Risk factors for oral cancer in women. *Cancer Res* 1989; 49: 2803–2806.
92. Kabat GC, Wynder EL. Type of alcoholic beverage and oral cancer. *Int J Cancer* 1989; 43: 190–194.
93. Merletti F, Boffetta P, Ciccone G, et al. Role of tobacco and alcoholic beverages in the etiology of cancer of the oral cavity/oropharynx in Torino, Italy. *Cancer Res* 1989; 49: 4919–4924.
94. Blot WJ, McLaughlin JK, Winn DM, et al. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res* 1988; 48: 3282–3287.
95. Sitas F, Urban M, Bradshaw D, et al. Tobacco attributable deaths in South Africa. *Tob Control* 2004; 13: 396–399.
96. Lewin F, Norell SE, Johansson H, et al. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck: a population-based case-referent study in Sweden. *Cancer* 1998; 82: 1367–1375.
97. Schlecht NF, Franco EL, Pintos J, et al. Interaction between tobacco and alcohol consumption and the risk of cancers of the upper aero-digestive tract in Brazil. *Am J Epidemiol* 1999; 50: 1129–1137.
98. Franceschi S, Levi F, La Vecchia C, et al. Comparison of the effect of smoking and alcohol drinking between oral and pharyngeal cancer. *Int J Cancer* 1999; 83: 1–4.
99. Zatonski W, Becher H, Lissowska J, et al. Tobacco, alcohol, and diet in the etiology of laryngeal cancer: a population-based case-control study. *Cancer Causes Control* 1991; 2: 3–10.
100. Bagnardi V, Blangiardo M, La Vecchia C, et al. Alcohol consumption and the risk of cancer: a meta-analysis. *Alcohol Res Health* 2001; 25: 263–270.
101. International Agency for Research on Cancer. Overall evaluations of carcinogenicity to humans. Group 1: carcinogenic to humans. Agents and groups of agents. Available at: <http://monographs.iarc.fr/eng/classification/crthgr01.php>. Accessed May 2009.
102. Salaspuro V, Salaspuro M. Synergistic effect of alcohol drinking and smoking on in vivo acetaldehyde concentration in saliva. *Int J Cancer* 2004; 111: 480–483.
103. Woo KS, Robinson JTC, Adams MR, et al. Differences in the effect of cigarette smoking on endothelial function in Chinese and white adults. *Ann Intern Med* 1997; 127: 372–375.
104. Liu BQ, Peto R, Chen ZM, et al. Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *BMJ* 1998; 317: 1411–1422.
105. Ishiguro S, Sasazuki S, Inoue M, et al. Effect of alcohol consumption, cigarette smoking and flushing response on esophageal cancer risk: a population-based cohort study (JPHC study). *Cancer Lett* 2009; 275: 240–246.
106. Vioque J, Barber X, Bolumar F, et al. Esophageal cancer risk by type of alcohol drinking and smoking: a case-control study in Spain. *BMC Cancer* 2008; 8: 221.
107. Ide R, Mizoue T, Fujino Y, et al. Cigarette smoking, alcohol drinking, and oral and pharyngeal cancer mortality in Japan. *Oral Dis* 2008; 14: 314–319.
108. Fan Y, Yuan JM, Wang R, et al. Alcohol, tobacco, and diet in relation to esophageal cancer: the Shanghai cohort study. *Nutr Cancer* 2008; 60: 354–363.
109. Stein L, Urban MI, Weber M, et al. Effects of tobacco smoking on cancer and cardiovascular disease in

- urban black South Africans. *Br J Cancer* 2008; 98: 1586–1592.
110. Bosetti C, Gallus S, Peto R, et al. Tobacco smoking, smoking cessation, and cumulative risk of upper aerodigestive tract cancers. *Am J Epidemiol* 2008; 167: 468–473.
111. Zendejdel K, Nyrén O, Luo J, et al. Risk of gastroesophageal cancer among smokers and users of Scandinavian moist snuff. *Int J Cancer* 2008; 122: 1095–1099.
112. Freedman ND, Abnet CC, Leitzmann MF, et al. Prospective investigation of the cigarette smoking-head and neck cancer association by sex. *Cancer* 2007; 110: 1593–1601.
113. Freedman ND, Abnet CC, Leitzmann MF, et al. A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am J Epidemiol* 2007; 165: 1424–1433.
114. Wang JM, Xu B, Rao JY, et al. Diet habits, alcohol drinking, tobacco smoking, green tea drinking, and the risk of esophageal squamous cell carcinoma in the Chinese population. *Eur J Gastroenterol Hepatol* 2007; 19: 171–176.
115. Hashibe M, Boffetta P, Janout V, et al. Esophageal cancer in Central and Eastern Europe: tobacco and alcohol. *Int J Cancer* 2007; 120: 1518–1522.
116. Ansary-Moghaddam A, Martiniuk A, Lam TH, et al. Smoking and the risk of upper aero digestive tract cancers for men and women in the Asia-Pacific region. *Int J Environ Res Public Health*. 2009; 6: 1358–1370.
117. Hashibe M, Brennan P, Chuang SC, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev* 2009; 18: 541–550.

SPECIAL FEATURE

Management of Stage D Heart Failure

Mark Harrison, MD, Anelechi Anyanwu, MD, and Sean P. Pinney, MD

Mount Sinai School of Medicine, New York, NY

Chronic heart failure is a major public health issue.¹ It affects more than 5.2 million Americans and accounts for more than 1 million hospitalizations annually. Heart failure is a deadly condition. After being diagnosed with heart failure, 1 in 5 patients will die within 1 year. Eighty percent of men and 70% of women less than 65 years of age who have heart failure will die within 8 years. Caring for patients with heart failure is expensive. In 2008, the annual estimated cost of caring for patients with heart failure was over 34 billion dollars.

In 2001, the American College of Cardiology and the American Heart Association reclassified the stages of heart failure, emphasizing that heart failure is a spectrum that encompasses patients with risk factors for the development of heart failure (stage A), those with structural heart disease who have never had symptoms of heart failure (stage B), those with symptomatic heart failure (stage C), and those with end-stage heart failure (stage D). The stages of heart failure progress linearly (Figure 1). For example, a survivor of a myocardial infarction who is initially free of heart failure symptoms is best classified as being in stage B. That individual will remain in that stage until heart failure symptoms develop, thereby advancing to stage C. Even if symptoms fully resolve, that individual will be identified as being stage C until heart failure progresses to its advanced and terminal stage, stage D.

The focus of this review is to highlight the management of patients with stage D heart failure.

Although the number of patients with asymptomatic heart failure (stages A and B) is about 4 times greater than the number of patients with symptomatic heart failure (stages C and D), the mortality rate for stage D heart failure patients is significantly higher, exceeding that of most cancers.² By the best estimates, there are currently about 250,000 patients with stage D heart failure, and their numbers are growing. Nonetheless, there is reason to be optimistic as the potential to improve the quality and extend the life expectancy of these individuals is greater now than at any other time in the past.

PROGNOSIS

Stage D heart failure is end-stage heart failure. In general, patients with a life expectancy of less than 6 to 12 months are best classified as belonging to stage D. Although one would desire to know when a heart failure patient transitions from stage C to D, the current level of diagnostic accuracy is lacking. Instead, heart failure cardiologists must rely on a combination of risk prediction models and clinical acumen. Two of the best prediction models are the Heart Failure Survival Score³ and the Seattle Heart Failure Model.⁴ The latter has the advantage of being offered as a web-based application (www.seattleheartfailure.com) and is derived solely from noninvasive clinical data. The Heart Failure Survival Score incorporates cardiopulmonary exercise testing and is the only model that has proven reproducibility over time.⁵ Some of the strongest predictors of poor survival include low serum sodium, an elevated resting heart rate, systemic hypotension, an inability to increase or a need to decrease the dose of an angiotensin-converting enzyme (ACE) inhibitor or beta-blocker, and a peak oxygen consumption of less than 12 mL/kg/minute.^{3,4} Recurrent hospital admission for the treatment of decompensated heart failure is not incorporated in these models but identifies

Address Correspondence to:

Sean P. Pinney

Advanced Heart Failure and
Cardiac Transplant Program
Mount Sinai Medical Center
New York, NY

Email: sean.pinney@mssm.edu

Figure 1. Stages in the Evolution of Heart Failure/Recommended Therapy by Stage



Fig 1. Stages of heart failure. **Abbreviations:** ACE, angiotensin-converting enzyme; FHx CM, family history of cardiomyopathy; HF, heart failure; IV, intravenous; LV, left ventricular; MI, myocardial infarction. Reprinted with permission from *Evaluation and Management of Chronic Heart Failure in the Adult ACC/AHA Pocket Guideline*.²⁶ Copyright 2002, American College of Cardiology/American Heart Association. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

a group of heart failure patients at greatest risk for recurrent admission and death.⁶

TREATMENT

The treatment options for stage D heart failure patients are limited (Figure 2). They include a continuous infusion of a positive inotrope, cardiac transplantation, implantation of a ventricular assist device (VAD) for destination therapy (DT), and palliative care. Cardiac transplantation and DT are able to extend and improve quality of life but are limited by a lack of access, and they consume a great deal of healthcare resources.⁷ By comparison, continuous outpatient support with inotropes (COSI) and hospice care offer comfort and relief from dyspnea but do not extend survival and, in some cases, may shorten life expectancy.⁸ Deciding between these 4 options requires an open dialogue between the heart failure cardiologist and patient so

that the patient understands what can be achieved, recognizing that for many, particularly those with advanced age or significant comorbidities the best route may be to choose improving quality of life.

Continuous Inotropic Support

Positive inotropic drugs, such as dobutamine and milrinone, have long been used to treat acute decompensated heart failure. The routine outpatient infusion of these medications is frowned upon because of the association with truncated survival. In stage D heart failure, these drugs are frequently employed to augment cardiac output, and occasionally patients are discharged home on a continuous infusion, usually while awaiting transplantation. Continuous outpatient support can offer palliation for those patients not eligible for cardiac transplantation or DT. This approach alleviates the symptoms of fatigue and dyspnea that accompany end-stage heart failure, but COSI does not prolong survival.⁸ More than half of patients on

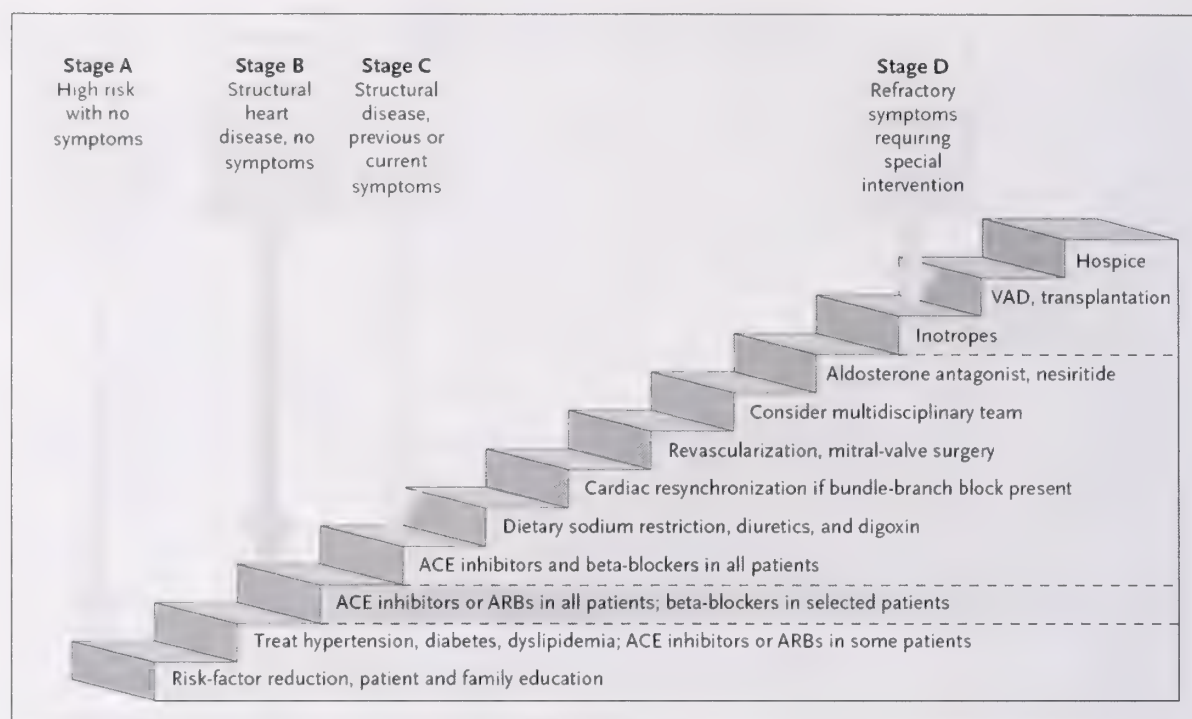


Fig 2. Therapies for the advancing stages of heart failure. **Abbreviations:** ACE, angiotensin-converting enzyme; ARB, angiotensin II receptor blocker; VAD, ventricular assist device. Adapted from the *New England Journal of Medicine*.²⁷ [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

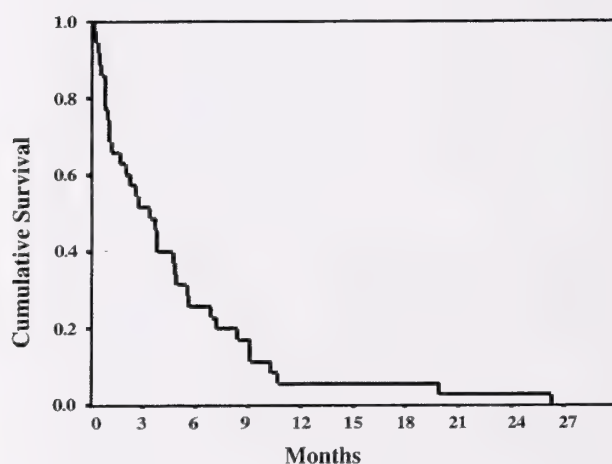


Fig 3. Probability of survival for advanced heart failure patients dependent on a continuous infusion of a positive inotrope. Adapted from the *Journal of Cardiac Failure*.⁸

COSI will die within the first 3 months, and only 5% to 10% will survive for 1 year (Figure 3). Although there remains a frequent need for hospitalization, many patients will be able to live at home until the end, maintaining some sense of independence.

Cardiac Transplantation

Cardiac transplantation remains the best choice for long-term survival for patients with stage D

heart failure. The first cardiac transplant was a xenotransplant performed by James Hardy in 1964 when he implanted a chimpanzee heart into a young girl (Figure 4). The honor of performing the first human allograft transplant belongs to Dr. Christiaan Barnard. In December 1967, he operated on a 53-year-old man in Cape Town, South Africa. The first transplant on American soil as well as the first pediatric transplant was performed days later by Dr. Adrian Kantrowitz in Brooklyn, NY. Dr. Norman Shumway, one of the most revered cardiac transplant surgeons because of his monumental contributions to the field, performed the fourth transplant in January 1968. His tireless research throughout much of the 1970s allowed the field to persevere when many were calling for a moratorium on cardiac transplantation. Many viewed the procedure as macabre, organs being harvested from brain-dead donors and implanted in recipients who often lived for merely weeks and frequently never left the hospital. Heart transplantation remained a medical curiosity until the introduction of cyclosporine in the 1980s. With this advance in immunosuppression, survival rates dramatically improved, and the number of heart transplants grew exponentially (Figure 5). Heart transplantation soon became the standard of care for advanced heart failure, offering 1-year survival rates in excess of 80%.

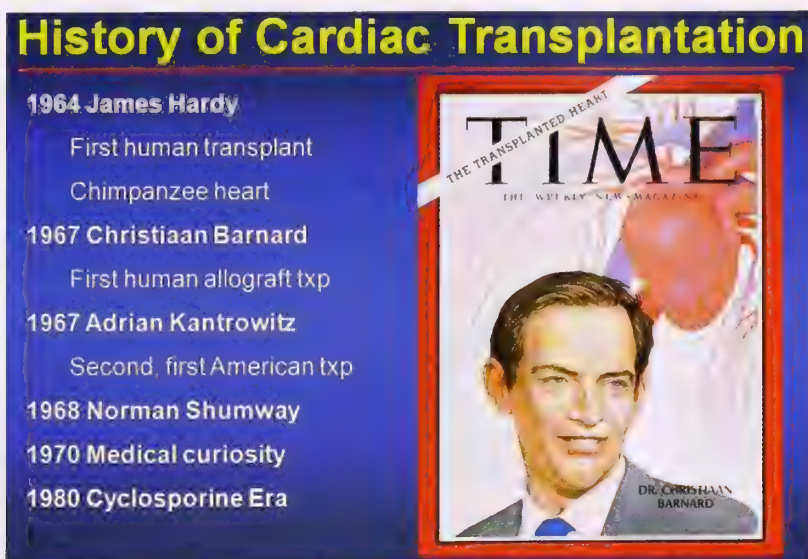


Fig 4. History of cardiac transplantation. **Abbreviation:** txp, transplant.

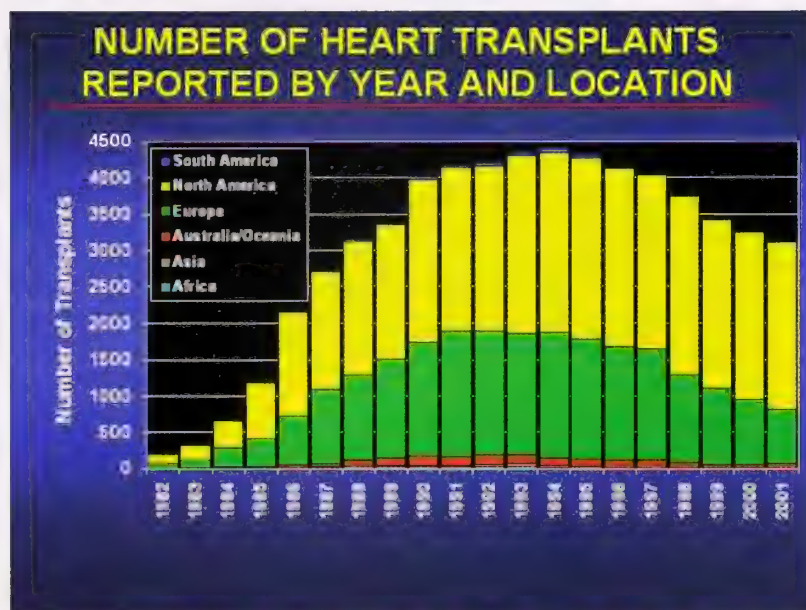


Fig 5. Number of heart transplants reported by year and location. Adapted from the International Society for Heart and Lung Transplantation.²⁸

The outcomes for heart transplant recipients continue to improve. One-year survival in the current era is between 85% and 90%, and median survival is nearly 10 years.⁹ Long-term survivors are now the expectation rather than the exception. Those patients who survive the first year, with its attendant risk from surgery, maximal immune suppression, and greatest risk of rejection, can expect a median survival of 13 years. Heart transplantation also improves quality of life. Ninety percent of recipients will have no limitation in their daily activities, and only 5% will

need some form of assistance. Total disability is rare (Figure 6).

Once again, there is a call from some in the heart failure community to perform a randomized clinical trial of heart transplantation.¹⁰ The basis for this call comes from the fact that survival is improving for chronic heart failure patients. The advent of neurohormonal antagonists, implantable cardiac defibrillators, and resynchronization devices (cardiac resynchronization therapy) has improved the survival of patients, including those currently waiting for

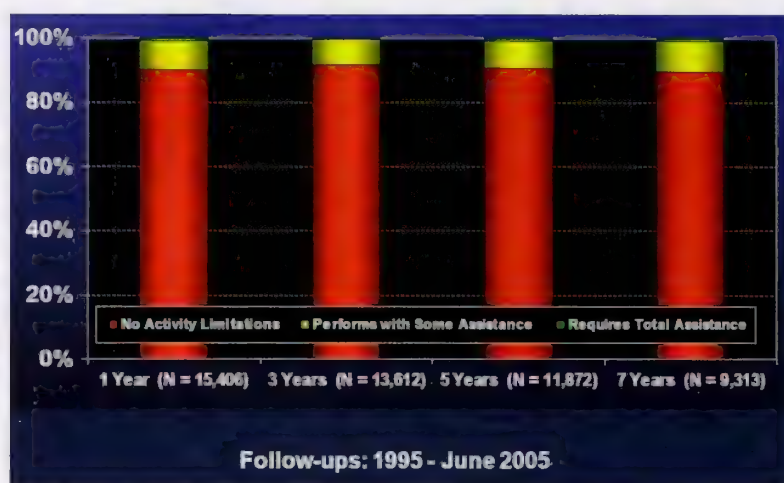


Fig 6. Functional status of surviving heart transplant recipients. Adapted from the International Society for Heart and Lung Transplantation.²⁸

transplantation.¹¹ Ambulatory patients listed for transplant but not requiring a positive inotrope infusion (status 2) now have a 1-year survival of nearly 90%. This appears to match or, in some cases, exceed expected posttransplantation survival. Although this comes from censored data, it appears that patients whose heart failure is not advanced to the point of requiring positive inotrope infusions may have little to gain by early transplantation. Because donation rates have not increased over the past 15 years, it is reasonable to delay listing for transplantation until such time as positive inotropes are required. In fact, a review of the listing status over the past 20 years highlights a trend away from listing status 2 patients and reserving listing for those with more advanced symptoms and shorter life expectancy.¹¹

Referral to a transplant center should be considered before patients are dependent on positive inotropes. This allows time for an adequate evaluation to determine if transplantation is necessary and provides patients with enough opportunity to make an informed decision. Although no strict referral criteria exist, there are several indicators that should prompt evaluation (Table 1). Most of these have been mentioned previously as predictors of poor heart failure survival. In addition to low serum sodium, intolerance of neurohormonal antagonists, and frequent heart failure hospitalization, one should consider evaluation for patients with New York Heart Association (NYHA) functional class III symptoms, a blood urea nitrogen level greater than 40 mg/dL, a diuretic requirement greater than 1.5 mg/kg/day furosemide or its equivalent, or a B-type natriuretic peptide or N-terminal pro-B-type natriuretic peptide level greater than 5 times the upper limit of normal. Once referred, heart failure patients consult

Table 1. Referral for Cardiac Transplantation.

NYHA III (unable to walk 2 blocks)
One or more heart failure hospitalizations in 6 months
Intolerant of ACE-I/ARB/beta blockers
Serum sodium <136 mmol/L
BUN >40 (excluding dehydration)
BNP or NT-proBNP >5 times normal
Diuretic requirement $\geq$ 1.5 mg/kg/day furosemide or equivalent

Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; BNP, B-type natriuretic peptide; BUN, blood urea nitrogen; NT-proBNP, N-terminal pro-B-type natriuretic peptide; NYHA, New York Heart Association.

with a cardiologist to ensure that reversible factors have been corrected, medical therapy has been maximized, and no significant comorbidities have been identified. A list of the exclusion criteria for our center is listed in Table 2.

The amount of time that one spends waiting for an acceptable donor is determined by listing status, body size, and blood type. In the United States, a new organ allocation protocol has been designed to distribute organs preferentially to patients with a higher listing status. Patients listed as status 1A have an expected life expectancy of 1 week. This is defined as the need for intensive care unit admission with the use of a pulmonary artery (Swan-Ganz) catheter to guide therapy. Status 1A patients must be receiving an infusion of a single, high-dose positive inotrope or a combination of 2 inotropes. Other criteria include the need for an intra-aortic blood pump, a VAD for less than 30 days, or mechanical ventilation. Status 1B patients are defined by the need for a continuous infusion of positive inotropes

Table 2. *Exclusion Criteria for Cardiac Transplantation.*

Age >70
Pulmonary hypertension: PVR >6 wood units
Chronic renal disease with CLcr <40 or serum creatinine >2.5 mg/dL
Active smoking, drug, or alcohol dependency
Diabetes with end-organ disease (retinopathy, nephropathy, and vasculopathy)
Severe obstructive/restrictive lung disease
Chronic liver disease
Pulmonary infarction within past 6 weeks
Major chronic disabling disease
Severe peripheral arterial or cerebrovascular disease
Obesity with BMI >35
Active or recent malignancy (within 5 years)
Psychological or social barriers that would preclude one's ability to reliably comply with medical therapy and follow-up examinations

Abbreviations: BMI, body mass index; CLcr, creatinine clearance; PVR, pulmonary vascular resistance.



Fig 7. The Heartmate XVE left ventricular assist device is a first-generation, pulsatile ventricular assist device with a textured titanium surface, which allows for the formation of a pseudointima that is resistant to thrombus formation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

or the presence of a VAD for more than 30 days. Status 2 patients are those listed for transplant but not requiring any of the above. The institution of this new allocation protocol has succeeded in distributing more organs to status 1A and 1B patients by sending donor hearts out of a given organ allocation region. The net result has been an appropriate redistribution to the neediest patients and a reduction in waiting list mortality. It has also created a situation in which one must be sicker in order to undergo transplantation. This is particularly true for patients with blood types associated with longer wait times, such as O and A, and those with a larger body habitus. As time spent on the waiting list increases, more patients are receiving a VAD to bridge the time from listing to identification of a suitable donor.

VENTRICULAR ASSIST DEVICES: BRIDGE TO TRANSPLANTATION

There are now 3 generations of VADs (Figures 7–9). First-generation VADs are pulsatile, positive-displacement pumps. They are by definition bulky as they must contain 60 to 80 mL of stroke volume to be displaced per ejection. All but one of the current Food and Drug Administration (FDA)–approved devices fall into this category. Examples of these first-generation pumps are the Heartmate XVE, the Thoratec paracorporeal VAD, the Thoratec implantable VAD, and the Novocar left ventricular assist system. Second-generation or axial-flow pumps generate a continuous flow through the use of a rotary impeller suspended in the blood stream by contact bearings. By comparison, these pumps are

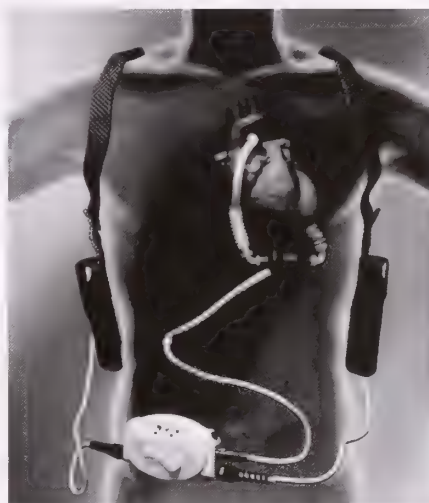


Fig 8. The Heartmate II is a second-generation, axial-flow device. It is smaller in size and quieter than first-generation pumps and is the only continuous-flow ventricular assist device currently approved for bridge to transplantation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

miniature, being about one-fifth of the size of the first-generation pumps. Only one, the Heartmate II, is FDA-approved for bridging to transplant. Other investigational second-generation devices include the Jarvik 2000 and Micromed DeBakey pumps. Third-generation VADs also create continuous flow but

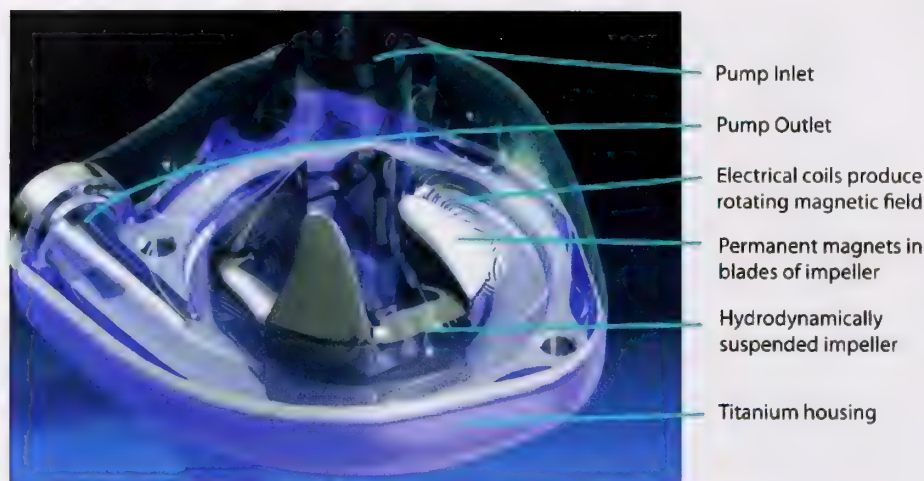


Fig 9. The Ventrassist is a third-generation ventricular assist device under clinical investigation for bridge to transplantation and destination therapy. It has a single impeller, which is hydrodynamically elevated and spins inside an electromagnetic field.

do so by the use of a single rotary impeller that is either magnetically or hydrodynamically suspended. By eliminating the need for contact bearings, these devices could, in theory, last forever and may be less prone to thrombus formation. Examples of these investigational devices include the Ventrassist, Heartware, and Duraheart pumps.

Many centers are turning away from the use of first-generation VADs, preferring instead to implant continuous-flow pumps. Despite their proven success in bridging patients to transplant,¹² the durability of these pulsatile VADs is limited by bearing wear and valvular insufficiency. Half of all Heartmate XVE pumps will experience a major device malfunction within 18 months of implantation.¹³ Pulsatile VADs are further disadvantaged by their bulky size and need for greater surgical dissection, which can complicate postoperative recovery. The need to ventilate the air that is displaced with each stroke requires a larger driveline that can be difficult to secure and is prone to infection. Moreover, patients frequently complain that these pumps are loud. On the other hand, the Heartmate XVE is the only VAD that does not require systemic anticoagulation. Its textured titanium surface allows for the formation of a pseudointima that is resistant to thrombosis. Unfortunately, this surface is also biologically active and can stimulate the production of alloantibodies, which in turn increase the risk of rejection following transplantation. Their presence effectively reduces the eligible donor pool and prolongs waiting times.

Continuous-flow VADs are quickly becoming the pumps of choice for bridge to transplantation. They are small, quiet, and durable. They have drivelines about the size of a telephone cord that are easy to

secure and rarely become infected. The Heartmate II is the first FDA-approved continuous-flow VAD. In the pivotal bridge-to-transplant trial, 80% of recipients either were alive on support or had received a transplant at 6 months of follow-up.¹⁴ Similar results were recently reported for the Ventrassist, a third-generation device (Figure 10). Despite their advantages over pulsatile pumps, these devices do pose some challenges. All continuous-flow VADs require systemic anticoagulation to prevent thrombus formation. Hemolysis is occasionally produced by the helical blades, which typically spin at 10,000 rpm for the Heartmate II and 2100 rpm for the Ventrassist. Higher spin rates produce greater rates of hemolysis, which in turn increase the chance of thrombus formation, particularly at the site of contact bearings where radial velocity is minimal. These VADs provide continuous, nonpulsatile flow. Although organ function does not seem to be negatively affected by the lack of pulsatility, continuous flow does make it difficult for clinicians to monitor patients because they may not be able to palpate a pulse or auscultate a blood pressure.

The small size and durability of these pumps are already producing remarkable results. Surgeons at Mount Sinai Medical Center reported the first implantation of a Jarvik 2000 through a median sternotomy without the use of cardiopulmonary bypass¹⁵ (Figure 11). The recipient of this device was removed from mechanical ventilation in less than 24 hours and sat up out of bed by the first postoperative day. Early reports such as these have been reinforced by observed shorter recovery times, a reduction in hospital length of stay, and improved survival. Patients occasionally feel so well with

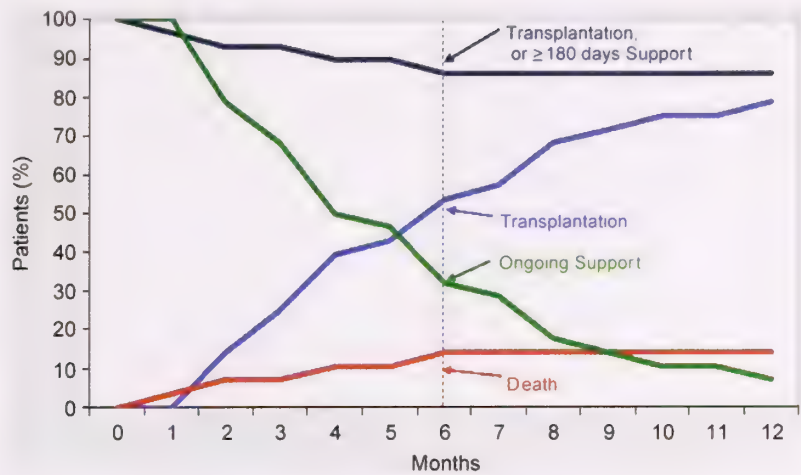


Fig 10. The competing outcomes of the pilot bridge-to-transplant trial of the Ventrassist device demonstrate excellent survival, with 86% of recipients having received a transplant or being alive on support 6 months after implantation.



Fig 11. First reported implantation of a Jarvik 2000 left ventricular assist device through a median sternotomy without the use of cardiopulmonary bypass (see Anyanwu *et al.*¹⁵ for details).

these second-generation and third-generation devices that they forego transplantation entirely. One well-documented patient, supported by a Jarvik 2000, survived more than 7 years, ultimately dying from noncardiac causes.¹⁶ A postmortem inspection of his device showed minimal to no bearing wear, suggesting that, in the absence of damage to the driveline or battery failure, this pump could reliably run for decades.

The timing of device implantation greatly influences clinical outcomes. Implantation of a VAD in the setting of chronic heart failure should ideally be a planned, semi-elective procedure. Emergency or bailout VAD implantation should be infrequent and reserved for acute onset cardiogenic shock. Early referral to specialized VAD centers is critical to obtaining successful outcomes.

The timing of VAD implantation is now guided by a stage D heart failure patient's INTERMACS (Interagency Registry for Mechanically Assisted Circulatory

Support) level (Table 3). The first registry report from INTERMACS, a multi-institutional database of outcomes for patients receiving FDA-approved VADs, has shown that postimplantation survival is adversely impacted by operations on patients with more advanced hemodynamic instability.¹⁷ For instance, if one were to wait for patients to develop critical

Table 3. INTERMACS Level.

- 1. Critical cardiogenic shock: “crash and burn”
- 2. Progressive decline: “sliding fast”
- 3. Stable but inotrope-dependent: “inotrope-dependent”
- 4. Recurrent advanced heart failure: “frequent flyer”
- 5. Exertion-intolerant: “housebound”
- 6. Exertion-limited: “walking wounded”
- 7. Advanced NYHA III

Abbreviations: INTERMACS, Interagency Registry for Mechanically Assisted Circulatory Support; NYHA, New York Heart Association.

cardiogenic shock before implanting a VAD, nearly 1 of 3 recipients would die by the third postoperative month. The same is true for nearly 1 of 5 patients who are “sliding fast” at the time of implantation. On the other hand, patients who are less acutely sick are more likely to survive by undergoing surgery before the advent of multiorgan failure accompanying cardiogenic shock. This is not to say that these patients are not sick. We know from clinical trials that the 1-year mortality rate for patients with NYHA IV heart failure is at least 75%.^{18,19} Although the actual timing of surgery remains a fluid decision, in general it is safe to conclude that earlier is better than later.

VENTRICULAR ASSIST DEVICES: DESTINATION THERAPY

Despite transplantation's clinical success, not all patients with stage D heart failure are eligible for this therapy. Even if they were, the rates of organ donation have not increased over the past 15 years. As a result, transplantation is clinically meaningful for the lucky few who receive a new heart but is epidemiologically trivial. Best estimates are that 250,000 patients have advanced heart failure, but only 1% will receive a transplant. VADs can help bridge this gap between supply and demand.

DT with a VAD can extend and improve quality of life for stage D heart failure patients. The landmark REMATCH (Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure) trial randomized transplant-ineligible patients either to receive the Heartmate XVE or to continue with optimal medical therapy.¹⁸ One-year survival with a VAD was nearly double that with optimal medical therapy. In addition, surveys of NYHA functional class and overall sense of well being improved with device therapy in comparison with patients in the medical therapy group. Although REMATCH is best viewed as proof of principle for DT, critics have pointed out that 1-year survival was only 50% for VAD recipients and that 3 of 4 patients had died by 2 years. Since REMATCH was published, a promising sign has emerged that familiarity with the device and improved management protocols have improved clinical outcomes.²⁰ Two-year survival now approaches 50% and has been accompanied by reductions in clinical adverse events and hospital length of stay. These improvements should lead to improved outcomes for VAD recipients and reduced costs for healthcare payers.⁷

Before becoming the treatment of choice for the majority of patients with advanced heart failure,

VAD therapy must further improve. Providers must know that they are choosing the right pump for the right patient at the right time. With their reduced size and improved durability, continuous-flow VADs are likely to become the pumps of choice for DT. Conclusive clinical trial results are as yet unknown, but the early termination of the Heartmate II DT trial at a prespecified interim analysis suggests that this continuous-flow VAD may indeed be superior to the Heartmate XVE. Trials of the Ventrassist pump remain ongoing but hold similar promise. The use of a preoperative risk score can now stratify potential recipients into low, medium, high, and very high risk categories for post-VAD mortality.²¹ Application of this multivariable score allows clinicians to reserve DT for those patients with an expected 1-year survival of at least 75%. For higher risk patients, an initial period of stabilization and improved nutrition prior to surgery may allow for eligibility at a later date as guided by this score.

VENTRICULAR ASSIST DEVICES: THE PROMISE OF RECOVERY

Until very recently, the process of left ventricular remodeling was believed to progress inexorably toward a globular, dysfunctional heart prone to mechanical and electrical failure. Medical therapy with an ACE inhibitor or beta-blocker could temporarily delay the remodeling process but could not reverse it. We now recognize that the volume and pressure unloading afforded by VADs can reverse some of the detrimental changes in the remodeled left ventricle. One month after VAD implantation, there is a beneficial reduction in left ventricular end-diastolic diameter and mass.²² Hypertrophied myocytes regress back toward a normal phenotype, and cardiac mechanics improve, particularly in those patients additionally treated with an ACE inhibitor.^{23,24} Although the net result of these changes is favorable, in most instances, they are insufficient to produce enough recovery of function to allow device removal and obviate transplantation. A recent report from Harefield, England has engendered hope that recovery may be attainable in more patients than originally thought possible.²⁵ In combination with an intensive medical regimen including the beta-2 agonist clenbuterol, 11 of 15 patients enrolled in their recovery protocol were able to safely undergo device removal. All enjoyed a dramatic improvement in their quality of life and freedom from heart failure symptoms. For recovery to become a reality for the majority of heart failure patients, we must better

understand myocyte biology and the signals involved in myocardial remodeling so as to be able to re-engineer the myocyte, allowing it to sustain function in the absence of mechanical support.

VENTRICULAR ASSIST DEVICES: FUTURE APPLICATIONS

Although VADs may allow for reverse remodeling of the left ventricle, they may also be able to provide a stable platform to re-engineer the heart. Much excitement has been engendered by the promise of using stem cells to regenerate myocardium, in essence regrowing the heart. In the future, patients may first receive a VAD to deliver a sufficient cardiac output to preserve end-organ function, which will be followed by an infusion of stem cells. A similar strategy may be employed by the combination of device therapy and gene therapy. A multicenter trial is currently underway to determine whether an adenovirus can successfully deliver to failing myocytes a gene encoding for sarco(endo)plasmic reticulum Ca^{2+} ATPase (SERCA), a protein pump involved in the handling of intracellular calcium. In failing hearts, SERCA is down-regulated and leads to impaired myocyte contractility. Although CUPID (Calcium Upregulation by Percutaneous Administration of Gene Therapy in Cardiac Disease) is not enrolling VAD patients, if it is successful, its results may further the promise of recovery for VAD recipients.

Perhaps the most exciting prospect for VADs is their becoming standard of care for patients with advanced heart failure. In order for this to occur, our concept of mechanical support needs to evolve. Our language and thinking are currently constrained by terms such as *destination* and *bridge* that payers employ. There is enough crossover between these labels to render them pointless. Some patients who receive a destination VAD improve to the point that they are eligible for transplantation, their device thus becoming a bridging VAD. Others who receive a bridging device develop complications that preclude transplant, so the VAD becomes DT. The labels distract us from the more important measurement of total patient outcomes. Consider too that patients with chronic kidney disease are evaluated for renal replacement when their disease has advanced to stage 4 or 5. These candidates are not offered dialysis as a bridge to transplantation, nor does one refer to hemodialysis as DT for those who are ineligible for transplant. This logic should be applied to advanced heart failure. Advanced heart failure itself—not postoperative expectations of transplantation—should be the indication needed for receiving a device.

CONCLUSION

The past few years have borne witness to the evolution of a new generation of VADs that have improved and extended the lives of patients with stage D heart failure. Although outcomes with cardiac transplantation continue to improve, the low rates of organ donation will continue to limit this therapy to a fortunate few. Bridging the gap between supply and demand will be VADs, which are poised to become the standard of care for patients with stage D heart failure. In order for that to occur, we as clinicians must first understand the long-term biology of VAD support as we try to unlock the mechanisms that lead to myocardial recovery. By March 2009, Medicare patients will be required to seek treatment at medical centers that have Joint Commission accreditation for VADs, a designation given to programs that have the multidisciplinary teams and experience needed to ensure the best perioperative results. As our experience and understanding of VADs grow, more patients with advanced heart failure will enjoy the benefits of these restorative devices.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Rosamond W, Flegal K, Furie K, et al. Heart disease and stroke statistics—2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 2008; 117: e25–e146.
2. Frigerio M, Oliva F, Turazza FM, et al. Prevention and management of chronic heart failure in management of asymptomatic patients. *Am J Cardiol* 2003; 91: 4–9.
3. Aaronson KD, Schwartz JS, Chen TM, et al. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation* 1997; 95: 2660–2667.
4. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation* 2006; 113: 1424–1433.
5. Lund L, Aaronson K, Mancini D. Validation of peak exercise oxygen consumption and the Heart Failure Survival Score for serial risk stratification in advanced heart failure. *Am J Cardiol* 2005; 95: 734–741.
6. Setoguchi S, Stevenson LW, Schneeweiss S. Repeated hospitalizations predict mortality in the community population with heart failure. *Am Heart J* 2007; 154: 260–266.
7. Hernandez AF, Shea AM, Milano CA, et al. Long-term outcomes and costs of ventricular assist devices among Medicare beneficiaries. *JAMA* 2008; 300: 2398–2406.

8. Hershberger R, Nauman D, Walker TL, et al. Care processes and clinical outcomes of continuous outpatient support with inotropes (COSI) in patients with refractory endstage heart failure. *J Card Fail* 2003; 9: 180–187.
9. 2007 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: Transplant Data 1997–2006. Rockville, MD: Health Resources and Services Administration; 2007.
10. Young JB. Heart failure's near dead and dying: reconsidering our heart transplant wait list scheme. *J Am Coll Cardiol* 2007; 50: 1291–1293.
11. Lietz K, Miller L. Improved survival of patients with end-stage heart failure listed for heart transplantation: analysis of Organ Procurement and Transplantation Network/U.S. United Network of Organ Sharing data, 1990 to 2005. *J Am Coll Cardiol* 2007; 50: 1282–1290.
12. Frazier OH, Rose EA, Oz MC, et al. Multicenter clinical evaluation of the HeartMate vented electric left ventricular assist system in patients awaiting heart transplantation. *J Thorac Cardiovasc Surg* 2001; 122: 1186–1195.
13. Pagani FD, Long JW, Dembitsky WP, et al. Improved mechanical reliability of the HeartMate XVE left ventricular assist system. *Ann Thorac Surg* 2006; 82: 1413–1418.
14. Miller LW, Pagani FD, Russell SD, et al. Use of a continuous-flow device in patients awaiting heart transplantation. *N Engl J Med* 2007; 357: 885–896.
15. Anyanwu AC, Fischer GW, Plotkina I, et al. Off-pump implant of the Jarvik 2000 ventricular assist device through median sternotomy. *Ann Thorac Surg* 2007; 84: 1405–1407.
16. Westaby S, Frazier OH, Banning A. Six years of continuous mechanical circulatory support. *N Engl J Med* 2006; 355: 325–327.
17. Kirklin J, Naftel DC, Stevenson LW, et al. INTERMACS database for durable devices for circulatory support: first annual report. *J Heart Lung Transplant* 2008; 27: 1065–1072.
18. Rose EA, Gelijns AC, Moskowitz AJ, et al., for the Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure (REMATCH) Study Group. Long-term use of a left ventricular assist device for end-stage heart failure. *N Engl J Med* 2001; 345: 1435–1443.
19. Packer M, Coats AJ, Fowler MB, et al. Effect of carvedilol on survival in severe chronic heart failure. *N Engl J Med* 2001; 344: 1651–1658.
20. Long JW, Healy AH, Rasmusson BY, et al. Improving outcomes with long-term “destination” therapy using left ventricular assist devices. *J Thorac Cardiovasc Surg* 2008; 135: 1353–1361.
21. Lietz K, Long JW, Kfoury AG, et al. Outcomes of left ventricular assist device implantation as destination therapy in the post-REMATCH era: implications for patient selection. *Circulation* 2007; 116: 497–505.
22. Maybaum S, Mancini D, Xydas S, et al. Cardiac improvement during mechanical circulatory support: a prospective multicenter study of the LVAD working group. *Circulation* 2007; 115: 2497–2505.
23. Klotz S, Foronjy RF, Dickstein ML, et al. Mechanical unloading during left ventricular assist device support increases left ventricular collagen cross-linking and myocardial stiffness. *Circulation* 2005; 112: 364–374.
24. Klotz S, Danser AH, Foronjy RF, et al. The impact of angiotensin-converting enzyme inhibitor therapy on the extracellular collagen matrix during left ventricular assist device support in patients with end-stage heart failure. *J Am Coll Cardiol* 2007; 49: 1166–1174.
25. Birks EJ, Tansley PD, Hardy J, et al. Left ventricular assist device and drug therapy for the reversal of heart failure. *N Engl J Med* 2006; 355: 1873–1884.
26. *Evaluation and Management of Chronic Heart Failure in the Adult ACC/AHA Pocket Guideline*. Washington, DC: American College of Cardiology/American Heart Association; 2002.
27. Jessup M, Brozena S. Heart failure. *N Engl J Med* 2003; 348: 2007–2018.
28. International Society for Heart and Lung Transplantation. Available at: <http://www.ishlt.org>. Accessed May 2009.

EVIDENCE-BASED MEDICINE

Accuracy of Computed Tomography Angiography for Coronary Artery Disease

Deborah Korenstein, MD, and Thomas McGinn, MD

Mount Sinai School of Medicine, New York, NY

This is the sixth in a series of articles in the *Mount Sinai Journal of Medicine* on evidence-based medicine, emphasizing applications of evidence-based medicine concepts to clinical practice. The articles will focus on the retrieval, interpretation, and application of evidence related to a current clinical problem. This is the first article in the series to focus on issues of diagnosis. The discussion will be based on the following case.

THE CASE

A 51-year-old woman with a history of mild hypertension and hyperlipidemia presents to her internist's office complaining of dyspnea on exertion. She states that her exercise tolerance used to be essentially unlimited but that now she notes shortness of breath after walking 5 or 6 blocks, which resolves with rest. Physical examination reveals an overweight woman in no distress and no physical examination abnormalities. Her electrocardiogram in the office is normal, except for a right bundle branch block, which is old.

The internist develops a differential diagnosis for this patient's shortness of breath, including deconditioning and coronary artery disease (CAD), and believes that the likelihood of CAD is about 35%,

which is high enough to warrant further work-up. The internist recommends a stress test to evaluate for CAD. The patient states that she has heard about the use of computed tomography (CT) scans to evaluate for heart disease and prefers to have that instead of a stress test if possible. She asks if a CT scan would be adequate to rule out heart disease in her case.

FOCUSED CLINICAL QUESTION

To determine the accuracy of CT scans for ruling out heart disease, the internist decides to review the relevant medical literature. He begins by formulating a focused question consisting of 4 elements: the patient (P), the intervention (I), the comparison (C), and the outcome (O), referred to as a PICO question.¹ The explicit inclusion of these 4 components allows for a more efficient and directed search of the literature and helps to clarify the exact problem of interest. The PICO format was designed for questions related to therapy but can be useful in diagnosis questions as well. In this case, the internist's question is: in women with a moderate likelihood of coronary disease (P), what is the accuracy of CT angiogram (I), in comparison with coronary angiography (C), in diagnosing significant CAD (O)?

SEARCH OF THE LITERATURE

The internist decides to go to PubMed to look for his answer because CT angiography is a relatively recent technology that may not yet be well described in textbooks. In PubMed, he goes to Clinical Queries. Clinical Queries is a search tool within PubMed that is designed to retrieve high-quality articles to answer

Address Correspondence to:

Deborah Korenstein
Mount Sinai School of Medicine
New York, NY
Email: deborah.korenstein@mountsinai.org

clinical questions. The user types relevant search terms and clicks a box to designate the type of question (etiology, therapy, diagnosis, prognosis, or clinical prediction guides) and the scope of the search (sensitive/broad versus specific/narrow). Additional terms are automatically added to each search to enable the retrieval of appropriate articles for that search type. Narrow or specific searches will retrieve fewer articles that are related closely to the search terms, and broad or sensitive searches will find more articles that may be more loosely related to the search terms. It is generally useful to begin with a specific search and move on to a sensitive search if there are too few retrieved articles.

The internist types “CT angiography ‘coronary artery disease’” into the search box, using quotation marks to ensure that those words are searched as one phrase and not as independent terms. He then clicks the diagnosis box and opts for a narrow/specific search, which is the default setting. The search yields 178 titles, which seem like too many to review. He clicks the limits button and limits the search to “humans,” “clinical trials,” and “English,” and this yields 17 articles. The second article on the list is titled “Diagnostic Performance of 64-Multidetector Row Coronary Computed Tomographic Angiography for Evaluation of Coronary Artery Stenosis in Individuals Without Known Coronary Disease: Results from the Prospective Multicenter ACCURACY (Assessment by Coronary Computed Tomographic Angiography of Individuals Undergoing Invasive Coronary Angiography) Trial.”² The internist decides to review that article.

Validity Criteria for Diagnosis Articles

In previous articles in this series, we have discussed validity issues in articles about treatment and prognosis. The issues around the validity of a diagnosis article are different and are often less intuitive to many physicians.

The major issues are as follows³:

- The clinicians faced diagnostic uncertainty. One of the major validity issues in studies evaluating diagnostic tests is related to the population being studied. The study population should reflect patients for whom the test would be useful in real practice in terms of disease severity, alternate diagnoses, practice setting, and demographic factors and in the sense of including patients who truly need a diagnostic test. This last issue is always important to consider because if the enrolled patients have a clear diagnosis, they reflect a population in which the test would not

be used in practice and in which the test might perform differently. These properties of the study population are known as **spectrum**. You can appreciate the importance of spectrum in studies of diagnostic tests if you consider an example of an inappropriate spectrum. Imagine a study evaluating a new diagnostic test for pulmonary embolism (PE) that enrolled patients from the intensive care unit, so all the patients who actually had PE had severe hypoxemia and hypotension. Even if the new diagnostic test appeared accurate in this population with very severe disease, the study would not offer useful information about how that test might perform in patients with small PEs without substantial hypoxemia.

In the ACCURACY trial, the investigators enrolled patients who had been referred for nonurgent invasive coronary angiography (ICA). When clinicians refer patients for ICA, they usually have a high suspicion of CAD and possibly of severe CAD. The patients in this study then may represent a population with severe disease and may differ substantially from patients with few risk factors and atypical chest pain in whom we might use this test in real practice. The clinicians caring for the study patients did face diagnostic uncertainty, but because the spectrum of patients in the study leaned toward severe disease, clinicians need to be careful to apply the results only within that spectrum of patients.

- The investigators used an appropriate **reference standard** to confirm the true diagnosis. When a new diagnostic test is being evaluated, it is important for the investigators to establish a reliable and unbiased method of determining the true diagnosis, which is known as the reference standard. The reference standard may be a single test, a series of tests, or a combination of methods including expert opinion or long-term follow-up. When the reference standard test involves subjective interpretation, studies of diagnostic test accuracy should use more than 1 interpreter of both the index test (the test being studied) and the reference standard test to minimize bias. Furthermore, the study should report how often the readers agree in their interpretation and should provide some systematic way of determining the official test result (by consensus or a third tiebreaker interpreter). The appropriate reference standard for CAD is ICA, although because ICA does not provide a functional assessment or coronary artery perfusion, it is an imperfect reference standard. This study did use ICA as the reference standard. However, despite the subjective nature of reading ICA, the study used

only 1 ICA reader. Furthermore, the researchers used computer software to quantify the degree of arterial stenosis, yet there is no evidence provided to support the validity of these computerized readings. This weakens the validity of the reference standard, but it is still reasonable.

- The reference standard and the CT scans were performed **independently**, meaning that the readers of each test were unaware of the results of the other test (ie, they were blinded). This is an important aspect of the validity of the trial because knowledge of the result of one of the tests could influence the interpretation of another test. This is a less important issue when test results are completely objective (eg, a number from a laboratory test), but when there is any subjectivity, as in this trial, blinding is important. In ACCURACY, CT and ICA were interpreted independently.
- The reference standard was performed on all patients, regardless of the result of the CT scan. This is an important aspect of validity because if the investigators do not perform the reference standard on subjects with negative tests, it is not possible to calculate the rate of false negatives or to calculate sensitivities or likelihood ratios (LRs). In studies such as this one with invasive reference standards, it is sometimes difficult to perform the reference standard on all patients. In the ACCURACY trial, all patients did in fact undergo both CT and ICA, the reference standard. The investigators accomplished this by enrolling patients who were already scheduled for ICA. The benefit of their approach is to facilitate every patient receiving both tests; the downside is the impact on spectrum, as discussed previously.

Although the internist has some concerns about spectrum and the reference standard, he determines that the study has overall good validity and decides to read the results.

Results

The study initially enrolled 245 patients but analyzed only the 230 patients who had completed both tests. Patient demographics are presented in the article. The mean age was 57 years, and 59% of the patients were male. The vast majority of the patients (88%) were Caucasian with a mean body mass index of 31.4. In terms of cardiac risk factors, 24% of the patients had diabetes, 67% had hypertension, 56% smoked, and 74% had a family history of CAD. In all, the population was at intermediate to high risk of CAD.

The prevalence of CAD in the population (with a cutoff of 50% stenosis on ICA used to define CAD) was 25%, which is surprisingly low given the spectrum of the enrolled patients.

There are several ways to look at the results in a study of diagnostic test accuracy. The major test properties that are most often reported are sensitivity and specificity. The sensitivity is the proportion of diseased patients who test positive, and the specificity is the proportion of nondiseased patients who test negative. Very sensitive tests are helpful in ruling out disease when negative and very specific tests are helpful in ruling in disease when positive. Sensitivity and specificity can be difficult to apply to patients; clinicians need to interpret positive or negative tests in the context of the patient at hand. Positive and negative predictive values seem to represent a more clinically relevant approach to the interpretation of diagnostic tests and are sometimes presented in studies of diagnostic test performance. Positive and negative predictive values, however, can be misleading because they reflect both the test performance characteristics and the prevalence of disease in the population studied. They do not allow for the application of the test result to a particular patient or population with different disease prevalence than that found in the population in the study.

The best method for applying diagnostic test results to a particular patient involves the use of likelihood ratios (LRs). LRs combine sensitivity and specificity and can be calculated for positive tests, negative tests, or different ranges of test results. LRs allow clinicians to determine the likelihood of disease in a particular patient, given a particular test result.

1. The following are the formulas for calculating LRs for positive (+) and negative (-) tests:

$$\text{LR+} = \text{Sensitivity} / 1 - \text{Specificity}$$

$$\text{LR-} = 1 - \text{Sensitivity} / \text{Specificity}$$

In the ACCURACY trial, the sensitivity of CT was 95%, and the specificity was 83%; this resulted in an LR+ of 5.6 and an LR- of 0.06. In general, an LR+ of >10 or an LR- of <0.1 is very helpful in making diagnoses. Because the LR- is <0.1 but the LR+ is not >10, we can say that a negative test is very useful for ruling out disease in general, but a positive test is less useful for confirming disease. Using a nomogram, we can determine the impact of a positive or negative test on the likelihood of disease, given different pretest probabilities, as shown in Table 1.

Table 1. *Relationship of the Pretest Probability and the Interpretation of Test Results.**

Patient	Pretest Probability	Posttest Probability for a Positive Test	Posttest Probability for a Negative Test
35-year-old man with a borderline lipid profile and atypical chest pain	5%	23%	<1%
48-year-old diabetic woman with intermittent chest pressure on exertion but good exercise tolerance	30%	71%	3%
65-year-old man with diabetes, hypertension, and sharp chest pain on exertion	60%	89%	8%

* For positive or negative CT Angio, using data from the ACCURACY trial.

APPLICABILITY: THE PHYSICIAN'S DECISION

The internist decides that because of issues around spectrum (including only patients who were going for ICA), the ACCORD study may not apply to his patient. However, he recognizes that CT angiography looks promising, especially for ruling out significant disease. Because a stress test would give prognostic information as well as diagnostic information, he encourages the patient to go for the stress test instead of the CT scan, although in discussing the issue with the patient, he reveals that a negative CT would probably rule out disease in patients like her.

CONCLUSION

This case illustrates several important points:

- Clinical Queries in PubMed can be used to find articles about the accuracy of diagnostic tests.
- When thinking about the validity of a study about a diagnostic test, one should look closely at the spectrum of patients included in the study and the reference standard used.
- LRs should be used to express the accuracy of diagnostic tests. LRs can be calculated for positive tests, negative tests, and ranges of test results and can be applied to different pretest probabilities to calculate posttest disease probabilities.

In future articles on evidence-based medicine, we will discuss other aspects of the evidence cycle and will delve further into studies of diagnostic accuracy. The complete cycle is shown here:

1. A case and dilemma.
2. A well-formulated clinical question.
3. A search of the literature to answer the question.
4. A critical appraisal of the literature.
5. The incorporation of the patient's values and preferences.
6. Finally, the application of the results to the care of the patient.

Each article in this series will focus on 1 or 2 specific areas, such as methods for searching the medical literature for meaningful data and the procedure for concealment in medical research. Although not meant to be exhaustive, this series will demonstrate how a physician can obtain reliable, evidence-based answers to specific clinical questions of importance to his or her patients.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12–A13.
2. Budoff MJ, Dowe D, Jollis JG, et al. Diagnostic performance of 64-multidetector row coronary computed tomographic angiography for evaluation of coronary artery stenosis in individuals without known coronary disease: results from the prospective multicenter ACCURACY (Assessment by Coronary Computed Tomographic Angiography of Individuals Undergoing Invasive Coronary Angiography) trial. *J Am Coll Cardiol* 2008; 52: 1724–1732.
3. Furukawa TA, Strauss S, Bucher HC, Guyatt G. Diagnostic tests. In: Guyatt G, Rennie D, Meade MO, Cook DJ, eds. *Users' Guides to the Medical Literature*. 2nd ed. New York, NY: McGraw-Hill; 2008: 419–438.

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 76, NO. 5

SEPTEMBER/OCTOBER 2009

Issue Theme

**Zoonotic Infections:
An Emerging Threat to
Human Health**

Theme Editors

**Stephanie Factor
Mary Klotman**

A Journal of Translational and Personalized Medicine



Surveillance and Control of Zoonotic Agents Prior to the Detection of Disease in Humans
The Bushmeat Trade: Increased Opportunities for the Transmission of Zoonotic Disease
Xenotransplantation, Xenogeneic Infections, Biotechnology and Public Health
Neuropathy Outbreak in Workers Exposed to Aerosolized Porcine Neural Tissue
The Significant but Understudied Impact of Pathogen Transmission from Humans to Animals
The Plague of Athens: Epidemiology and Palaeopathology
Malaria Management in the 20TH Century: The Rockefeller Foundation Program

Special Features

Paradigm Shift in the Pathophysiology of Post-menopausal and Thyrotoxic Osteoporosis
Controversies in the Management of Alcoholic Liver Disease
The Treatment of Primary Spinal Tumors

 **WILEY-BLACKWELL**



Giant
Resources...
Purebred
Content

Wiley InterScience® Reference Works

Sample Title List

Encyclopedia of Life Sciences
Encyclopedia of Statistical Sciences
Kirk-Othmer Encyclopedia of Chemical Technology
Sax's Dangerous Properties of Industrial Materials
Ullmann's Encyclopedia of Industrial Chemistry
Van Nostrand's Encyclopedia of Chemistry

Capabilities

Full text HTML and PDF (for some)
Reference linking
Citation download

Access

Single article or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Today's sophisticated researcher demands fast and convenient online access to the giants of scientific reference.

From *Kirk-Othmer* to the *Encyclopedia of Life Sciences*,

Wiley InterScience® Reference Works deliver the biggest names in scientific research literature to your desktop, right when you need them. Encompassing subjects ranging from chemistry, life sciences, and medicine to psychology, engineering, and physics, these multi-volume encyclopedias, handbooks, and dictionaries are enhanced by powerful functionality that accelerates your path to discovery. With **Wiley InterScience**, you can be sure to **Discover Something Great.**



WILEY
InterScience®
DISCOVER SOMETHING GREAT



1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editor

David H. Adams

Associate Editors

Stuart Aaronson
Erwin Bottinger
Dennis S. Charney
Valentin Fuster
Wayne Goodman
Jonathan Halperin
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Michael L. Marin
Thomas G. McGinn
Ira Nash
Maria New
Rosamond Rhodes
Hugh Sampson
Mary Sano
Jeffrey H. Silverstein
Albert Siu
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Theme Editors

Stephanie H. Factor, MD, MPH
Assistant Professor of Medicine
Division of Infectious Diseases
Mount Sinai School of Medicine

Mary E. Klotman, MD
Professor of Medicine and Microbiology
Co-Dir., Global Health & Emerging Pathogens Inst.
Mount Sinai School of Medicine

Editorial Director

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com



While examining a chicken's mouth, this 16-year old boy wears a mask to protect himself against avian influenza. He knows that this disease can be transmitted to humans directly from birds or through intermediate hosts and that the major route of transmission is respiratory secretions. This is an example of how an improved understanding of zoonotic diseases can help protect human health.

Future *MSJM* Themes and Theme Editors

- Alzheimer's disease (Nikos Robakis)
- Endovascular surgery (Michael Marin)
- Obstetrics and gynecology (Rhoda Sperleng)
- Genes, Race, Ethnicity and the Environment (Gary Butts)
- Drug discovery in academic settings (Ravi Iyengar)
- Children's health and the environment (Philip Landrigan)

Error Correction

In the *MSJM* issue 76:4 on Innovations in Medical Education, the authors of the article entitled "The Student-run Health Clinic: A Novel Arena to Educate Medical Students on Systems-based Practice" were listed in the wrong order. The correct order is the following: Andrea Wershof Schwartz, Jeremy S. Abramson, Israel Wojnowich, Robert Accordino, Edward J. Ronan, Mary R. Rifkin.

©2009 Mount Sinai School of Medicine. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: \$95.00 in North America, \$120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS

VOL 76, NO 5

Theme Introduction

- 419** Zoonotic Infections: An Emerging Threat to Human Health
Stephanie H. Factor, Mary E. Klotman

Theme Articles

- 421** Surveillance and Control of Zoonotic Agents Prior to the Detection of Disease in Humans
James E. Childs, Elizabeth R. Gordon
- 429** The Bushmeat Trade: Increased Opportunities for the Transmission of Zoonotic Disease
William B. Karesh, Eric Noble
- 435** Xenotransplantation, Xenogeneic infections, Biotechnology and Public Health
Louisa E. Chapman
- 442** Outbreak of Progressive Inflammatory Neuropathy in Workers following Exposure to Aerosolized Porcine Neural Tissue
Tracy M. DeAngelis, Liang Shen
- 448** The Significant but Understudied Impact of Pathogen Transmission from Humans to Animals
Jonathan H. Epstein, Joni Price
- 456** The Plague of Athens: Epidemiology and Palaeopathology
Robert J. Littman
- 468** Historical Perspectives on Malaria, Focusing on the Rockefeller Anti-Malaria Strategy in the 20th Century
Darwin H. Stapleton

Special Features

- 474** A Paradigm Shift in the Pathophysiology of Post-menopausal and Thyrotoxic Osteoporosis
Mone Zaidi, Jameel Iqbal, Harry C Blair, Alberta Zallone, Terry Davies, Li Sun
- 484** Controversies In the Management of Alcoholic Liver Disease
Hui-Hui Tan, Sharad Virmani, Paul Martin
- 499** The Treatment of Primary Vertebral Tumors
Harshpal Singh, Scott A. Meyer, M.D, Arthur L. Jenkins III

Complimentary Online Access for Institutions!

MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Editor-in-Chief: **Penny A. Asbell, MD, FACS, MBA**
Mount Sinai School of Medicine, New York, NY

Continuously published since 1934 and included in the *Index Medicus* from the very first volume, the **Mount Sinai Journal of Medicine (MSJM)** has a distinguished career as a general medical journal for a worldwide audience.

Building upon this tradition of education across medical specialties, **MSJM** was re-launched in 2007 as a journal of translational and personalized medicine, two disciplines in the forefront of modern healthcare. **MSJM** features articles on the evolving nature of clinical care, its close linkage to bio-medical research and the resulting opportunities for enhanced patient care.

Dedicated to improving communication among all participants in medicine today, **MSJM** is an excellent source of up-to-date medical information for healthcare clinicians, educators, researchers and policy makers worldwide. **MSJM** will inspire the creative thinking needed to create new ways to understand, prevent, and treat human disease.

Each issue includes several quality, in-depth review articles related to a single theme and several others on subjects of diverse clinical interest. All articles are written by renowned authorities in their fields, from both Mount Sinai and elsewhere, under the supervision of Guest Editors.

Published Issues

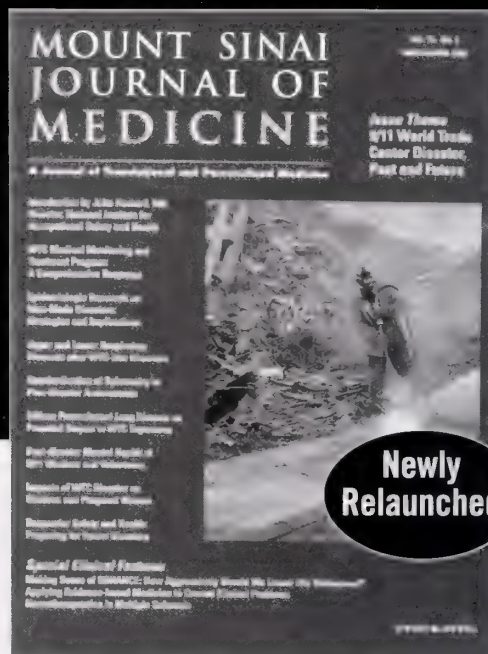
- Quality of Care and Racial Health Disparities
Mark R. Chassin (Vol. 75/1)
- 9/11 World Trade Center Disaster: Past and Future
Robin Herbert and Jacqueline Moline (Vol. 75/2)
- Anxiety and Depression: Leading Edge of Therapy
Sanjay J. Mathew and Dennis S. Charney (Vol. 75/3)
- Type 1 Diabetes Mellitus: New Understanding of Etiology, New Approaches to Management
Ronald Tamler (Vol. 75/4)
- Hospitalists: New Team in Forefront of Hospital Quality and Education
Andrew Dunn (Vol. 75/5)
- Diversity in Academic Medicine: A Call to Action
Gary C. Butts (Vol. 75/6)
- Advanced Lung Disease: Comorbidities and Challenges
Maria L. Padilla and Virginia R. Litle (Vol. 76/1)

Forthcoming Issues

- Traumatic Brain Injury (Vol. 76/2)
- Type 2 Diabetes Mellitus (Vol. 76/3)
- Medical Education (Vol. 76/4)
- Zoonotic Infections (Vol. 76/5)
- Drug Discovery in Academic Setting (Vol. 76/6)



WILEY-
BLACKWELL



COMPLIMENTARY ONLINE ACCESS FOR INSTITUTIONS!

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

Institutional online and print pricing for the **Mount Sinai Journal of Medicine** will be set in 2009, subsequent to introductory complimentary opt-in access during 2008 and 2009. Pricing will be based on the overall demand for the Journal and the size and type of the subscribing institution (e.g. academic, government, hospital, or corporate). Institutions that receive complimentary online access during 2009 will have archival access in subsequent years according to Wiley-Blackwell's licensing terms and conditions.

How to Receive Complimentary Online Access for Institutions?

Ask your librarian to e-mail optinaccess@wiley.co.uk or visit www.interscience.wiley.com/msjregister for more details and to register for complimentary online access to the **Mount Sinai Journal of Medicine**.

Library Recommendation Form

To _____

Date _____

Requested by: _____

Dept. _____

I have reviewed the **Mount Sinai Journal of Medicine** and urge you to add a complimentary 2009 online subscription to our subscription list. Please contact me should you have any questions.

THEME INTRODUCTION

Zoonotic Infections: An Emerging Threat to Human Health

Stephanie H. Factor, MD, MPH, and Mary E. Klotman, MD

Division of Infectious Diseases, Mount Sinai School of Medicine, New York, NY

In the spring of 2007, Dr. Paul Klotman, Chair of the Department of Medicine at Mount Sinai School of Medicine, proposed that a new course be developed for the Master of Public Health Program. He envisioned a course that used a global health perspective to introduce students to the public health issues associated with emerging zoonotic infections. The idea was his alone; an appropriate textbook could not be found. His request was in the context of a marked shift in the focus and expansion of international efforts at Mount Sinai through the creation of the Global Health and Emerging Pathogens Institute.

Although the term *global health* is frequently used, it has not been well defined. In order to standardize language, the Consortium of Universities for Global Health Executive Board in June 2009 defined global health as follows:

An area for study, research, and practice that places a priority on improving health and achieving equity in health for all people worldwide. Global health emphasizes transnational health issues, determinants, and solutions; involves many disciplines within and beyond the health sciences and promotes interdisciplinary collaboration; and is a synthesis of population-based prevention with individual-level clinical care.¹

They compared this to *international health*, which most often is used to refer to various health experiences abroad, often in resource-limited

countries, historically a focus among the trainees at Mount Sinai School of Medicine. Given the dramatic growth of basic and translational sciences at the institution that could affect health worldwide, the time had come to broaden the agenda.

Although global health encompasses a lot of areas, the emphasis on emerging pathogens highlights unique strengths at Mount Sinai in viral diseases. Of the over 350 emerging infectious diseases from 1940 to 2004, over 60% are considered zoonoses, human pathogens that originate from an animal source. The majority of zoonoses have arisen from wildlife as opposed to domestic animals. The percentage of zoonoses arising from wildlife has increased over this time.² Factors promoting the emergence of disease include ecological changes, such as those due to agricultural or economic development or anomalies in climate; human demographic changes and behavior; travel and commerce; technology and industry; microbial adaptation and change; and breakdown of public health measures. The interplay of these factors leading to recent worldwide outbreaks is best exemplified by 3 viral diseases: the entrance of human immunodeficiency virus 1 into humans, the rapid spread of severe acute respiratory syndrome, and the recent spread of Pandemic Influenza (H1N1) 09.

The Institute of Global Health and Emerging Pathogens at Mount Sinai has chosen a unique focus on the interplay between ecosystems and the emergence of zoonoses. To enhance the training of future investigators in this area, a course was developed for public health students. Rather than providing an in-depth review of zoonotic pathogens, as is often presented to medical students, or exhaustively describing modes of transmission of disease from animals to humans, the course involves a panel of experts who share their experiences in transnational determinants, treatments, and prevention of zoonotic diseases on the population and individual levels.

Address Correspondence to:

Stephanie H. Factor

Division of Infectious Diseases
Mount Sinai School of Medicine
New York, NY

Email: stephanie.factor@
exchange.mssm.edu

This issue is a collection of papers presented at or inspired by the 2008 class “MPH 010: Zoonoses: An Emerging Public Health Issue” given for the Mount Sinai School of Medicine’s Master of Public Health Program and cosponsored by the Global Health and Emerging Pathogens Institute. The authors come from many disciplines: history, classics, public health, human medicine, veterinary medicine, and epidemiology. Their work spans the globe. They address both known and recently identified diseases. They are advocates of both human and nonhuman animal health. They speak of the transmission of disease by age-old practices, such as hunting, and the transmission of disease by “of the moment” exposures, such as xenotransplantation and highly mechanized food preparation. This volume is not a

compilation of all that is important in understanding the emergence of zoonotic disease. It is a look at disease with a whole new set of eyes.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Koplan JP, Bond TC, Merson MH, et al. Towards a common definition of global health. *Lancet* 2009; 373: 1993–1995.
2. Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature* 2008; 451: 990–993.

Surveillance and Control of Zoonotic Agents Prior to Disease Detection in Humans

James E. Childs, ScD,¹ and Elizabeth R. Gordon, DVM, MD²

¹Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT

²Department of Emergency Medicine, Flint River Hospital, Montezuma, GA

OUTLINE

FROM DETECTION TO INTERVENTION: HUMAN-BASED APPROACHES
LIMITATIONS TO HUMAN-BASED
INTERVENTION PROGRAMS FOR CONTROLLING ZONOSSES
OBSTACLES TO ANIMAL-BASED SURVEILLANCE
QUARANTINE, ISOLATION, AND LEGISLATION
CULLING OF DOMESTIC LIVESTOCK AND POULTRY FOR ZONOSIS CONTROL
CULLING OF VECTORS AND WILDLIFE FOR ZONOSIS CONTROL
WILDLIFE AND INTERMEDIATE HOST
VACCINATION FOR ZONOSIS CONTROL
OTHER METHODS FOR ZONOSIS CONTROL
LATE-BREAKING DEVELOPMENTS

ABSTRACT

The majority of newly emerging diseases are zoonoses caused by pathogens transmitted directly or indirectly through arthropod vectors to humans. Transmission chains leading to human infection frequently involve intermediate vertebrate hosts, including wildlife and domestic animals. Animal-based surveillance of domestic and wild animals for zoonotic pathogens is a global challenge. Until

recently, there has been no scientific, social, or political consensus that animal-based surveillance for zoonotic pathogens merits significant infrastructural investment, other than the fledgling efforts with avian influenza. National institutions charged with strategic planning for emerging diseases or intentional releases of zoonotic agents emphasize improving diagnostic capabilities for detecting human infections, modifying the immune status of human or domestic animals through vaccines, producing better antiviral or antibacterial drugs, and enhancing human-based surveillance as an early warning system. With the exception of human vaccination, these anthropocentric approaches target post-spillover events, and none of these avenues of research will reduce the risk of additional emergences of pathogens from wildlife. Novel schemes for preventing spillover of human pathogens from animal reservoir hosts can spring only from an understanding of the ecological context and biological interactions that result in zoonotic disease emergence. Although the benefits derived from investments to improve surveillance and knowledge of zoonotic pathogens circulating among wildlife reservoir populations are uncertain, our experience with human immunodeficiency virus and the pandemic influenza inform us of the outcomes that we can expect by relying on detection of post-spillover events among sentinel humans. ***Mt Sinai J Med* 76:421–428, 2009.** © 2009 Mount Sinai School of Medicine

Key Words: influenza virus, surveillance, wildlife, zoonotic disease.

Address Correspondence to:

James E. Childs

Department of Epidemiology
and Public Health

Yale University School of
Medicine

New Haven, CT

Email: jamesechilds@comcast.net

Surveillance for zoonotic pathogens is largely based on detecting illness or infection in *Homo sapiens*; humans often serve as the sentinel species for zoonotic agents maintained in transmission cycles in which they typically serve as a dead-end host. A variety human-based surveillance systems have successfully captured the emergence of novel zoonotic disease or elucidated the incidence of

Humans often serve as the sentinel species for zoonotic agents maintained in transmission cycles in which they typically serve as a dead-end host.

previously known, but incompletely understood zoonotic diseases.¹

With a few exceptions, systematic and mandated national surveillance for zoonotic diseases among wildlife, as opposed to domestic animals and livestock, falls through the cracks of both veterinary and human health practices. Reviews of animal health monitoring systems mention wildlife disease surveillance only in passing.² The difficulties inherent to establishing wildlife-based pathogen/disease surveillance and the planning for integrating such data streams with human disease surveillance are daunting challenges.

Once a zoonotic disease is characterized, formal, human-based systematic surveillance efforts can be initiated at the state or national level in countries possessing the requisite infrastructure. Most national surveillance programs coordinated through the Centers for Disease Control and Prevention focus on the systematic collection of data on human disease; rabies reporting is the sole exception. Rabies in both humans and domestic or wild animals is a condition legally mandated as reportable to the Centers for Disease Control and Prevention from all states and territories. The clear association between the risk of human infection and the occurrence of rabies in mammalian hosts, in addition to animal testing as a major criterion for initiating postexposure treatment, dictates that surveillance be conducted at the level of the animal host.³

Surveillance for viruses transmitted from wildlife reservoir hosts to domestic poultry and livestock, such as avian and swine influenza A, is primarily conducted through the US Department of Agriculture, which also conducts mandated surveillance for zoonotic infections of livestock, such as bovine spongiform encephalopathy (BSE), anthrax, and bovine tuberculosis (TB). Several influenza subtypes infecting wildlife and domestic animals cause disease in humans.⁴ The current pandemic caused by the novel H1N1 subtype⁵ and the threat of pandemic disease posed by the H5N1 subtype⁶ have opened the door to enhancing surveillance efforts through partnerships with extragovernmental agencies such as the Wildlife Conservation Society.⁷

In the United States, individual states or institutions often take the lead in monitoring wildlife

diseases, such as surveilling for West Nile virus by the detection of the virus in dead birds submitted to state laboratories⁸ and tracking the spread of transmissible spongiform encephalopathy associated with elk and white-tailed deer.⁹ In California and Florida, sentinel bird flocks have been used for decades to monitor enzootic arboviral activity.⁸ Long-term data collected from these flocks have proven effective in elucidating trends in the occurrence of western equine encephalomyelitis, St. Louis encephalitis, and eastern equine encephalomyelitis virus linked to climatic and local weather patterns.^{10,11}

Studies of zoonotic pathogen transmission among reservoir hosts and to humans have been established in response to specific instances of emerging zoonotic disease. After the discovery of hantavirus pulmonary syndrome in the southwestern United States in 1993, multiyear, replicated studies were established in several states to learn more about Sin Nombre virus maintenance in natural populations of the reservoir host, the deer mouse (*Peromyscus maniculatus*),^{12,13} and its spillover to humans. Following the outbreak of human monkeypox in the United States,¹⁴ several states, in collaboration with the Centers for Disease Control and Prevention, initiated studies of local populations of indigenous North American rodents to discern if ongoing transmission was occurring in sites near animal holding facilities housing African rodents. Although African rodents imported for the pet trade were the original source of the virus, human monkeypox cases were linked to exposure to a native species of prairie dogs, popular as pets, which had been housed in close association with the African species.¹⁵

FROM DETECTION TO INTERVENTION: HUMAN-BASED APPROACHES

National institutions charged with strategic planning for emerging diseases or intentional releases of zoonotic agents have emphasized improving diagnostic capabilities for detecting human infections, modifying the immune status of human or domestic animals through vaccines, producing better antiviral or antibacterial drugs, and enhancing human-based surveillance as an early warning system.^{16,17} The current fixation on human-based interventions and surveillance is the legacy of past successes. Landmark achievements for zoonotic disease prevention include the development of vaccines for yellow fever and rabies, and additional vaccines for zoonotic diseases exist or are being developed. With the exception of

extensive human vaccination, each of these activities targets post-spillover events, and none will reduce the risk of additional emergences of zoonotic pathogens from wildlife.

The current fixation on human-based interventions and surveillance is the legacy of past successes. . . . With the exception of extensive human vaccination, each of these activities targets post-spillover events, and none will reduce the risk of additional emergences of zoonotic pathogens from wildlife.

LIMITATIONS TO HUMAN-BASED INTERVENTION PROGRAMS FOR CONTROLLING ZOONoses

As traditional approaches to zoonotic disease control have largely ignored the ecology of wildlife/pathogen maintenance and interactions at the animal-human interface required for pathogen emergence, only limited advances have been made in approaches designed to interrupt zoonotic pathogen transmission prior to human spillover.

Traditional public health prevention measures, including case isolation, contact tracing and quarantine of exposed persons, banning of public gatherings, and limiting access to international travel, proved highly effective in controlling the spread of certain zoonotic diseases with pandemic potential. The effectiveness of these methods is well illustrated by the successful control of severe acute respiratory disease syndrome.¹⁸ However, increasing social distance and quarantining ill or exposed persons proved ineffective in controlling the pandemic spread of H1N1. Unlike severe acute respiratory disease syndrome, for which the onset of clinical disease in humans is coincident with the onset of infectiousness, the aerosol transmissibility of influenza virus precedes clinical signs in humans by 24 hours.¹⁹ These differences dictate that the development of influenza A vaccines should remain the highest priority. Antiviral stockpiles are insufficient to treat large populations, and preexisting strains of H1N1 viruses have rapidly developed resistance to oseltamivir (brand name Tamiflu), which is considered one of the front-line medications.^{20,21}

OBSTACLES TO ANIMAL-BASED SURVEILLANCE

An ideal goal for a public health program to prevent zoonotic diseases would be to abrogate or greatly reduce the transmission and maintenance of pathogens within the animal reservoir prior to any human emergence. However, the full force of scientific creativity and funding has been focused on anthropocentric fixes rather than on understanding and then targeting potentially vulnerable processes leading to the emergence of human disease. Designing and implementing wildlife-based surveillance and targeted, pre-spillover interventions are not appropriate for most zoonotic diseases as the costs will be prohibitive. However,

The need to integrate animal and human surveillance activities is imperative.

when pandemic disease has occurred or remains a possibility, as is the case with influenza A subtypes and human immunodeficiency virus, integrated surveillance efforts targeting wildlife and humans may provide information for early and targeted interventions. The need to integrate animal and human surveillance activities is imperative.²²

Animal-based surveillance is a process inherently different from human-based surveillance. With the exception of systematic surveillance efforts targeting livestock and poultry and coordinated by the Center for Animal Health Surveillance of the US Department of Agriculture, no formal sampling methodology, outside of that designed for limited studies,^{12,13} exists for estimating animal population sizes and infection prevalence at the regional or continental level. Wildlife population estimates at the local or regional scale are few, and continental-based efforts, such as the national waterfowl surveys,²³ were designed for purposes other than disease surveillance. The lack of population data at most scales precludes the estimation of basic epidemiological parameters, including rates such as incidence or mortality. However, targeted studies at local scales have provided information that, when suitably leveraged through the use of statistical modeling and with geographical information system technology, can provide risk maps at regional levels, as exemplified by Lyme disease, plague, and hantavirus pulmonary syndrome.^{24,25}

When death or profound behavioral changes accompany disease occurring among wildlife (eg, rabies infection in mammals and West Nile virus

infection of certain bird species), surveillance efforts can capitalize on public participation in reporting suspicious events.^{26,27} However, many zoonotic pathogens establish asymptomatic infections among individuals of the reservoir host species (eg, Sin Nombre virus infection in deer mice and many influenza A viral infections among wild waterfowl). Wildlife-based surveillance for these pathogens is necessarily difficult and expensive, as it requires blind sampling of apparently healthy individuals.

QUARANTINE, ISOLATION, AND LEGISLATION

Quarantining foreign animals or animal products arriving in a country has a long history.²⁸ The major international regulation governing the international trade and movement of animal and plant products, the Agreement on the Application of Sanitary and Phytosanitary Measures, was established in 1995 along with the World Trade Organization.²⁹ The Agreement on the Application of Sanitary and Phytosanitary Measures was designed to facilitate animal trade while reducing the risk of exporting diseased animals or animal products, and its standards are set by the Office International des Epizooties. Of the current bans in effect, those governing the import of cattle and cattle products from nations reporting BSE and regulating the import of poultry and poultry products from countries with enzootic avian influenza³⁰ are most widely known. However, other examples include a federal law prohibiting the interstate translocation of potential rabies virus reservoir species, which was imposed after an outbreak of animal rabies in Florida initiated by the import of infected coyotes from Texas,³¹ and a Chinese law prohibiting trade in certain carnivore species, which was imposed in the wake of the severe acute respiratory disease syndrome outbreak. Export bans and culling of potentially infected food animals have enormous economic and political impacts and confront the stakeholders in a global economy with unprecedented new risks.^{32,33}

CULLING OF DOMESTIC LIVESTOCK AND POULTRY FOR ZONOSIS CONTROL

Control of emerging zoonotic agents circulating among domestic poultry, livestock, and companion animals is often targeted at specific infected subpopulations or demographic cohorts. Strategies

for national control of BSE have included destroying entire herds of cattle after identification of an index case, birth-cohort culling of animals born during a specific time interval, and maternal culling to prevent vertical transmission.³⁴ Culling of domestic poultry is the primary control for pathogenic influenza A subtypes.³⁵ In 2004, over 20 million chickens were killed in 8 nations of Southeast Asia in response to the looming threat of a pandemic spread of the influenza H5N1 virus.³⁶ Swine culling, following the 1997 outbreak of Nipah virus in Malaysia, had major economic consequences caused by trade bans and the direct killing of more than 1 million pigs.³⁷ Nipah virus has since reemerged in Malaysia, precipitating new rounds of culling.

CULLING OF VECTORS AND WILDLIFE FOR ZONOSIS CONTROL

The most widespread approach to zoonosis control has been the wide-scale culling of individuals of the wildlife reservoir or the intermediate hosts of pathogens.³⁸ The most common example of culling is the use of insecticides to control arthropod vectors and nuisance populations of mosquitoes. Culling of wildlife reservoir host populations has

The most widespread approach to zoonosis control has been the wide-scale culling of individuals of the wildlife reservoir or the intermediate hosts of pathogens.

been adopted or is planned to curtail transmission of several viral and bacterial pathogens to humans or domesticated livestock, although the record of population reduction as an effective prevention strategy to limit human spillover is mixed.³⁸

As an example, the reduction of specific reservoir host populations has been a cornerstone of rabies control. Although globally domestic dogs remain the primary host involved in rabies virus transmission to humans, nations that have effectively controlled canine rabies have placed their sights on wildlife reservoirs. In Europe and North America, culling programs have targeted red foxes and (in North America) raccoon and skunk populations.^{39,40} However, the development and dissemination of an oral rabies vaccine (ORV) for terrestrial wildlife has largely supplanted culling for rabies control in Europe and North America (as discussed later). In Central and South America, rabies transmitted to cattle by

vampire bats inflicts significant economic losses,⁴¹ and humans are also victims of these hematophagous animals. Vampire bat population control efforts use anticoagulants topically applied to bats or systemically introduced into cattle, and they have been linked to die-offs of nontarget bat species.⁴¹

A prominent and controversial use of culling has targeted wild badgers and brush-tailed possums (New Zealand) in an attempt to control bovine TB by removing wildlife reservoirs for the bacterium. The removal of 20,000 badgers in England from 1975 to 1997 failed to curb bovine TB spread among cattle,⁴² although reports from Ireland indicated that badger culling reduced the incidence of TB in cattle herds.⁴³ Vaccination of badgers against TB is now being investigated as a part of an integrated control program targeting cattle herds in specific locations.

WILDLIFE AND INTERMEDIATE HOST VACCINATION FOR ZOOONOSIS CONTROL

The second major approach to zoonotic pathogen control is the vaccination of populations of wildlife or intermediate hosts serving as reservoirs. Wildlife

The second major approach to zoonotic pathogen control is the vaccination of populations of wildlife or intermediate hosts serving as reservoirs.

vaccination is limited to a few species and pathogens, although new vaccines are under development.

Vaccination of livestock or poultry serving as intermediates for pathogen transmission to humans or at risk themselves has been used against Japanese encephalitis virus infection among swine, which serve as an amplifying host,⁴⁴ and against influenza A virus subtypes infecting chickens.⁴⁵ Although an effective ORV has been developed for domestic dogs, parental vaccination remains the standard.

Millions of doses of ORV have been disseminated to control red fox rabies in Europe and raccoon and red fox rabies in North America.^{46,47} With the advent of ORV, red fox rabies has been eliminated or greatly reduced in western Europe and southern Canada. In the United States, ORV has reduced levels of enzootic rabies transmission among raccoons in Ohio, West Virginia, and Pennsylvania and among gray foxes in Texas.⁴⁸ The rationale for raccoon immunization is the prevention of the westerly spread

of raccoon-variant rabies from enzootic regions in the eastern United States by the construction of an immune barrier.⁴⁹ In addition to their cost, ORV interventions demand a long-term public commitment.⁵⁰ However, traditional methods of rabies prevention are also costly,⁵¹ and ORV should be viewed as a supplementary approach suited to specific circumstances.

OTHER METHODS FOR ZOOONOSIS CONTROL

In the future, immune contraception could be used as an ethically acceptable alternative to the culling

In the future, immune contraception could be used as an ethically acceptable alternative to the culling of wildlife species critical to zoonotic pathogen maintenance.

of wildlife species critical to zoonotic pathogen maintenance. There are also practical limitations to culling; for instance, culling by hunting can be difficult when the target species lives in densely populated human environments. An example of a sensitive situation that limits population control options is the case of white-tailed deer, which readily populate suburban areas. Deer serve as the keystone host for several tick species that transmit *Borrelia burgdorferi*, the causative agent of Lyme disease,⁵² and *Ehrlichia* spp., which cause monocytic ehrlichiosis and a form of granulocytic ehrlichiosis.⁵³ Population control of white-tailed deer⁵⁴ has been shown to be an effective method to reduce the number of ticks and incidence of these diseases.

The genetic manipulation of arthropods or their endosymbionts to render vector species refractory to pathogen transmission is an area of intense research.^{55,56} In the future, such methods may be an important component in the control of vector-borne diseases; however, the only successful method involving the release of modified vectors has involved the use of sterile males to control veterinary and agricultural pests.

LATE-BREAKING DEVELOPMENTS

Since the drafting of this report, the US Agency for International Development has published a request

for collaboration on an agreement to carry out proposals outlined for 2 new and major programs, PREDICT and RESPOND:

The goal of PREDICT⁵⁷ is to establish a global early warning system for disease emergence that is capable of detecting, tracking and predicting the emergence of new infectious diseases in high-risk wildlife (e.g., bats, rodents, and non-human primates) that could pose a major threat to human health.

The goal of RESPOND⁵⁸ is to improve the capacity of countries in high risk areas to respond to outbreaks of emergent diseases that pose a serious risk to human health.

Additional information on these novel efforts to integrate human-based and wildlife-based surveillance and intervention activities can be found at the US Agency for International Development Web site.⁵⁹

SUMMARY

Most emerging diseases are zoonotic and/or vector-borne. The universe of opportunities available to zoonotic pathogens to interact with and explore an unprecedented (and evolutionarily unanticipated) variety of potential new niches has dramatically expanded with the introduction of exotic domesticated species and the intentional or accidental movements of wildlife and arthropod vector species. Although the detection and recognition of zoonotic diseases appearing among human sentinels have been the public health standard for surveillance and for directing interventions for disease control, there is a vast, virtually untapped opportunity to improve these efforts. The integration of human, veterinary, and agricultural medicine should result in earlier warning of disease emergence and offer approaches for intervention beyond the traditional anthropocentric measures.

ACKNOWLEDGMENT

The authors thank 2 anonymous reviewers who provided comments and corrections that improved this report.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Childs JE. Pre-spillover prevention of emerging zoonotic diseases: what are the targets and what are the tools? *Curr Top Microbiol Immunol* 2007; 315: 389–443.
2. Salman MD. *Animal Disease Surveillance and Survey Systems: Methods and Applications*. Ames, IA: Iowa State Press; 2003; 1–222.
3. Childs JE, Krebs JW, Smith JS. Public health surveillance and the molecular epidemiology of rabies. In: Leitner T, ed. *The Molecular Epidemiology of Human Viruses*. Dordrecht, The Netherlands: Kluwer Academic; 2002; 273–312.
4. Fouchier RA, Schneeberger PM, Rozendaal FW, et al. Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc Natl Acad Sci U S A* 2004; 101: 1356–1361.
5. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 2009; 459: 931–939.
6. Melidou A, Gioula G, Exindari M, et al. Influenza A(H5N1): an overview of the current situation. *Euro Surveill* 2009; 14: 19216.
7. Wildlife conservation Society [Internet]. New York: Wildlife conservation Society; c199502002. Avian Influenza (H5N1) [cited 2009 June 13]. Available from: <http://www.wcs.org/conservation-challenges/wildlife-health/disease-investigations/avian-influenza-h5n1.aspx>.
8. Eidson M, Kramer L, Stone W, et al. Dead bird surveillance as an early warning system for West Nile virus. *Emerg Infect Dis* 2001; 7: 631–635.
9. Williams ES, Miller MW. Transmissible spongiform encephalopathies in non-domestic animals: origin, transmission and risk factors. *Rev Sci Tech* 2003; 22: 145–156.
10. Reeves WC. *Epidemiology and Control of Mosquito-Borne Arboviruses in California, 1943–1987*. Sacramento, CA: California Mosquito and Vector Control Association; 1990; 1–508.
11. Day JF, Stark LM. Transmission patterns of St. Louis encephalitis and eastern equine encephalitis viruses in Florida: 1978–1993. *J Med Entomol* 1996; 33: 132–139.
12. Mills JN, Yates TL, Ksiazek TG, et al. Long-term studies of hantavirus reservoir populations in the southwestern United States: rationale, potential, and methods. *Emerg Infect Dis* 1999; 5: 95–101.
13. Mills JN, Ksiazek TG, Peters CJ, Childs JE. Long-term studies of hantavirus reservoir populations in the southwestern United States: a synthesis. *Emerg Infect Dis* 1999; 5: 135–142.
14. Centers for Disease Control and Prevention. Multistate outbreak of monkeypox—Illinois, Indiana, and Wisconsin, 2003. *MMWR Morb Mortal Wkly Rep* 2003; 52: 537–540.
15. Cunha BE. Monkeypox in the United States: an occupational health look at the first cases. *AAOHN J* 2004; 52: 164–168.
16. Fauchi AS. *NIAID Strategic Plan for Biodefense Research*. Bethesda, MD: National Institutes of Health; 2002; 1–12. NIH publication 03–5306.
17. Centers for Disease Control and Prevention. *Preventing Emerging Infectious Diseases: A Strategy for the 21st*

- Century*. Atlanta, GA: US Department of Health and Human Services; 1998; 1–75.
18. Zhong NS, Zeng GQ. Our strategies for fighting severe acute respiratory syndrome (SARS). *Am J Respir Crit Care Med* 2003; 168: 7–9.
 19. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci USA* 2004; 101: 6146–6151.
 20. Mills CE, Robins JM, Lipsitch M. Transmissibility of 1918 pandemic influenza. *Nature* 2004; 432: 904–906.
 21. Moscona A. Global transmission of oseltamivir-resistant influenza. *N Engl J Med* 2009; 360: 953–956.
 22. King LJ, Marano N, Hughes JM. New partnerships between animal health services and public health agencies. *Rev Sci Tech* 2004; 23: 717–725.
 23. Butler WI Jr, Stehn RA, Balogh GR. GIS for mapping waterfowl density and distribution from aerial surveys. *Wildl Soc Bull* 1995; 23: 140–147.
 24. Ogden NH, St Onge L, Barker IK, et al. Risk maps for range expansion of the Lyme disease vector, *Ixodes scapularis*, in Canada now and with climate change. *Int J Health Geogr* 2008; 7: 24.
 25. Eisen RJ, Glass GE, Eisen L, et al. A spatial model of shared risk for plague and hantavirus pulmonary syndrome in the southwestern United States. *Am J Trop Med Hyg* 2007; 77: 999–1004.
 26. Vekslor A, Eidson M, Zurbenko I. Assessment of methods for prediction of human West Nile virus (WNV) disease from WNV-infected dead birds. *Emerg Themes Epidemiol* 2009; 6: 4.
 27. Childs JE, Krebs JW, Real LA, Gordon ER. Animal-based national surveillance for zoonotic disease: quality, limitations, and implications of a model system for monitoring rabies. *Prev Vet Med* 2007; 78: 246–261.
 28. Gensini GF, Yacoub MH, Conti AA. The concept of quarantine in history: from plague to SARS. *J Infect* 2004; 49: 257–261.
 29. Zepeda C, Salman M, Thiermann A, et al. The role of veterinary epidemiology and veterinary services in complying with the World Trade Organization SPS agreement. *Prev Vet Med* 2005; 67: 125–140.
 30. Hall C. Impact of avian influenza on U.S. poultry trade relations-2002: H5 or H7 low pathogenic avian influenza. *Ann N Y Acad Sci* 2004; 1026: 47–53.
 31. Centers for Disease Control and Prevention. Translocation of coyote rabies—Florida, 1994. *MMWR Morb Mortal Wkly Rep* 1995; 44: 580–587.
 32. James A. The state of veterinary epidemiology and economics. *Prev Vet Med* 2005; 67: 91–99.
 33. von Overbeck J. Insurance and epidemics: SARS, West Nile virus and Nipah virus. *J Insur Med* 2003; 35: 165–173.
 34. European Commission [Internet]. European Communities, 1995–2009. Opinion of the Scientific Steering Committee on BSE-related culling in Cattle; 2000 Sept 15 [cited 2009 June 13]. Available from: http://ec.europa.eu/food/fs/sc/ssc/out138_en.pdf.
 35. Watts J. Asian nations step up action to curb spread of avian influenza. Outbreak is spreading at an unprecedented speed, WHO says, and nowhere can be considered safe. *Lancet* 2004; 363: 373.
 36. Abbott A, Pearson H. Fear of human pandemic grows as bird flu sweeps through Asia. *Nature* 2004; 427: 472–473.
 37. Uppal PK. Emergence of Nipah virus in Malaysia. *Ann N Y Acad Sci* 2000; 916: 354–357.
 38. Wobeser G. Disease management strategies for wildlife. *Rev Sci Tech* 2002; 21: 159–178.
 39. Debbie JG. Rabies control of terrestrial wildlife by population reduction. In: Baer GM, ed. *The Natural History of Rabies*. Boca Raton, FL: CRC Press; 1991; 477–484.
 40. Muller J. The effect of fox reduction on the occurrence of rabies. Observations from two outbreaks of rabies in Denmark. *Bull Off Int Epizoot* 1971; 75: 763–776.
 41. Kobayashi Y, Ogawa A, Sato G, et al. Geographical distribution of vampire bat-related cattle rabies in Brazil. *J Vet Med Sci* 2006; 68: 1097–1100.
 42. Delahay RJ, Wilson GJ, Smith GC, Cheeseman CL. Vaccinating badgers (*Meles meles*) against *Mycobacterium bovis*: the ecological considerations. *Vet J* 2003; 166: 43–51.
 43. Griffin JM, Williams DH, Kelly GE, et al. The impact of badger removal on the control of tuberculosis in cattle herds in Ireland. *Prev Vet Med* 2005; 67: 237–266.
 44. Daniels TJ, Williams DT, Mackenzie JS. Japanese encephalitis virus. In: Morrilla A, Yoon KJ, Zimmerman JJ, eds. *Trends in Emerging Viral Infections of Swine*. Ames, IA: Iowa State Press; 2002; 249–263.
 45. Lee CW, Senne DA, Suarez DL. Effect of vaccine use in the evolution of Mexican lineage H5N2 avian influenza virus. *J Virol* 2004; 78: 8372–8381.
 46. Pastoret PP, Brochier B. Epidemiology and control of fox rabies in Europe. *Vaccine* 1999; 17: 1750–1754.
 47. Hanlon CA, Niezgoda M, Morrill P, Rupprecht CE. Oral efficacy of an attenuated rabies virus vaccine in skunks and raccoons. *J Wildl Dis* 2002; 38: 420–427.
 48. Steelman HG, Henke SE, Moore GM. Bait delivery for oral rabies vaccine to gray foxes. *J Wildl Dis* 2000; 36: 744–751.
 49. Slate D, Rupprecht CE, Rooney JA, et al. Status of oral rabies vaccination in wild carnivores in the United States. *Virus Res* 2005; 111: 68–76.
 50. Kemere PK, Liddel MK, Evangelou P, et al. Economic analysis of a large scale oral vaccination program to control raccoon rabies. *Proc NWRC Spec Symp* 2002; 3: 106–116.
 51. Gordon ER, Krebs JW, Rupprecht CR, et al. Persistence of elevated rabies prevention costs following post-epizootic declines in rates of rabies among raccoons (*Procyon lotor*). *Prev Vet Med* 2005; 68: 195–222.
 52. Barbour AG. Fall and rise of Lyme disease and other *Ixodes* tick-borne infections in North America and Europe. *Br Med Bull* 1998; 54: 647–658.
 53. Paddock CD, Yabsley MJ. Ecological havoc, the rise of white-tailed deer, and the emergence of *Amblyomma americanum*-associated zoonoses in the United States. *Curr Top Microbiol Immunol* 2007; 315: 289–324.
 54. Wilson ML, Levine JF, Spielman A. Effect of deer reduction on abundance of deer tick (*Ixodes dammini*). *Yale J Biol Med* 1984; 57: 697–705.
 55. Rasgon JL, Styer LM, Scott TW. *Wolbachia*-induced mortality as a mechanism to modulate pathogen transmission by vector arthropods. *J Med Entomol* 2003; 40: 125–132.

56. Olson KE, Adelman ZN, Travanty EA, et al. Developing arbovirus resistance in mosquitoes. *Insect Biochem Mol Biol* 2002; 32: 1333–1343.
57. Grants.gov [Internet]. Washington, DC: U.S. Department of Health and Human Services. Grants.gov. Avian and Pandemic Influenza and Zoonotic Disease Program PREDICT [cited 2009 June 13]. Available from: <http://www.grants.gov/search/search.do?mode=VIEW&oppld=47965>
58. Grants.gov [Internet]. Washington, DC: U.S. Department of Health and Human Services Grants.gov. Avian and Pandemic Influenza and Zoonotic Disease Program RESPOND [cited 2009 June 13]. Available from: <http://www.grants.gov/search/search.do?mode=VIEW&oppld=46813>
59. US Agency for International Development. Available at: <http://www.usaid.gov/contact.html>. Accessed June 2009.

The Bushmeat Trade: Increased Opportunities for Transmission of Zoonotic Disease

William B. Karesh, DVM¹ and Eric Noble, BA²

¹Global Health Program, Wildlife Conservation Society, Bronx, NY

²Mount Sinai School of Medicine, New York, NY

OUTLINE

BUSHMEAT TRADE

CULTURAL, POLITICAL, AND ECONOMIC FACTORS
PROMOTING EXPANSION OF THE BUSHMEAT TRADE
IN AFRICA

Commercial Logging Armed Conflict

HUNTING FOR CONSENSUS: SOLUTIONS TO THE
BUSHMEAT CRISIS

Law Enforcement Agricultural and Other Development Surveillance

the emergence of zoonoses require a multidisciplinary approach. *Mt Sinai J Med* 76:429–434, 2009. © 2009 Mount Sinai School of Medicine

Key Words: bushmeat, emerging infectious disease, wildlife, zoonotic disease.

ABSTRACT

Bushmeat is a term that refers to the use of wild animals, ranging from cane rats to gorillas, for food. The term typically refers to the practice in forests of Africa. The bushmeat trade is an example of an anthropogenic factor that provides opportunities for the transmission of diseases from wildlife to humans. The expansion of the bushmeat trade over the past 2 decades has provided a venue for the emergence of zoonotic diseases by providing an increased opportunity for the transmission of organisms known to cause disease and organisms with an unknown impact on humans. Because the bushmeat trade is embedded in a complex cultural, political, and economic context, efforts to prevent

Infectious diseases account for 16.2% of global deaths,¹ and as highlighted in *Emerging Infections: Microbial Threats to Health in the United States*,² several infectious diseases are emerging as even greater threats to public health globally. Emerging infectious diseases are diseases that are either newly recognized in the human population or increasing in incidence. Of the diseases identified as emerging, approximately 75% are zoonotic³; that is, they can be naturally transmitted between humans and animals. Over the last few years, researchers have tried to understand the reasons underlying the emergence of zoonotic diseases. Most pathogens infecting animals fail to make the jump into humans, and only 33% of zoonotic species are known to be transmissible between humans.³ Why then are zoonotic diseases emerging more frequently in the human population?

Three steps have been identified as necessary for a zoonotic disease to become a pandemic.⁴ These steps are also useful in understanding the increasing incidence of zoonotic diseases. First, a pathogen must be successfully transmitted between a wild reservoir and humans. Second, the pathogen must be directly or indirectly transmitted between humans. Finally, the pathogen must be able to move from a local epidemic into the global population.⁴ At each of these 3 stages, there are factors that can increase the probability of a pathogen advancing to the next stage. Anthropogenic changes that affect the contact between animals and humans, such as hunting and butchering, can significantly increase the probability of the first step.⁵ Anthropogenic changes that affect the contact between persons, such as global travel,

Address Correspondence to:

William B. Karesh
Global Health Program,
Wildlife Conservation Society,
Bronx, NY
Email: wkaresh@twcs.org

urbanization, and increases in the human population, can significantly increase the probability of the third step.

BUSHMEAT TRADE

Bushmeat is a term referring to the use of wild animals, ranging from cane rats to gorillas, for food. The term typically refers to the practice in forests of Africa. Cultural, political, and economic factors have led to an increase in the bushmeat trade over the past 2 decades and thus increased the opportunities for the successful transmission of disease from a wild reservoir to a host, the first step in emergence.

Bushmeat is a term referring to the use of wild animals, ranging from cane rats to gorillas, for food. Cultural, political, and economic factors have led to an increase in the bushmeat trade over the past 2 decades and thus increased the opportunities for the successful transmission of disease from a wild reservoir to a host, the first step in emergence.

The drivers and mechanisms used to provide wildlife for food, medicine, pets, novelties, and other products are highly complex and highly varied by region, use, and wildlife species. The consumption of wild animals specifically for food occurs worldwide. In the forests of Central Africa, where the ability to raise domestic animals is limited, bushmeat serves as a major source of protein. Although the consumption of wild animals for food occurs in remote areas in South America, for much of the continent, domestic animals, fish, and processed products are widely available and limit dependence on wildlife as a major protein source.

The hunting of wildlife for products in addition to food has had a tremendous impact on wildlife resources in some areas of the world. In Asia, for example, there is a rapidly growing international trade in wildlife as exotic food and medicine that is driven by a rising demand in urban areas and southern China. As a result, many areas in southeast Asia that serve as a source of this trade are virtually devoid of wildlife.

In all places where it is practiced, the wildlife trade has site-specific complex drivers and mechanisms that defy easy solutions. We caution the reader that this is not an extensive review of the wildlife or bushmeat trade. To demonstrate the public health risks of the bushmeat trade and the difficulty of designing interventions that promote safety and health, we present the specific drivers and mechanisms of the bushmeat trade in Africa.

CULTURAL, POLITICAL, AND ECONOMIC FACTORS PROMOTING EXPANSION OF THE BUSHMEAT TRADE IN AFRICA

Contact with the blood and body fluids of animals during hunting and butchering is a well-known route of transmission of disease-causing organisms from animals to humans. Tularemia, for example, is also known as rabbit disease because it affects rabbit hunters. The expanded bushmeat trade, therefore, provides the opportunity for increased transmission of organisms known to cause disease and organisms with an unknown impact on humans. For example, a recent study has shown that the bushmeat trade provides opportunities for the transmission of human retroviruses.⁵ The emergence of the human immunodeficiency virus (HIV) in the 1980s, the retrovirus that causes acquired immune deficiency syndrome, is likely a result of the bushmeat trade.⁶

Over the past 2 decades, the intensity and spatial extent of the bushmeat trade have increased substantially, and this has allowed for greater contact between humans and animals.⁷ Estimates indicate that in the Congo Basin alone, 1 to 3.4 million tons of bushmeat (~2 billion pounds) are extracted annually.⁵ Many factors have contributed to this increase, particularly the expansion of commercial logging and the disruption of local trading practices due to armed conflict.

Commercial Logging

Commercial logging has transformed hunting from a primarily subsistence activity into a commercial enterprise by improving transportation, populating the frontier zone, and increasing bushmeat demand.⁷ The transportation infrastructure provided by the logging activity brings roads, trucks, guns, and hungry workers and their families into forest areas once far less accessible. Historically, hunting activities radiated in a circular fashion from isolated villages. Roads provide an increased number of points at which hunting activities can commence, "changing the pattern of human

contact from a circular pattern to a banded pattern surrounding developed roads, increasing the area in which hunting can be conducted with economic returns."⁴ Commercial logging also brings many more persons into the frontier zones, places where national laws against hunting apes and selling their flesh often go unenforced.⁸

In addition to providing an increased opportunity for bushmeat hunting, commercial logging appears to create an increased demand for bushmeat. Auzel and Wilkie,⁹ in their study of workers in a logging camp in northern Congo, found an increase in demand for bushmeat among logging workers due to disposable incomes and few other dietary options. Seventy-six percent of their meals contained bushmeat. Household bushmeat consumption also increased in nearby villages, from 39% of meals in unaffected villages to 49% of meals in villages servicing the logging camp, presumably because of the increase in hunting income.

Armed Conflict

Armed conflict has led to an increase in the bushmeat trade by disrupting the customary, local control measures. Specifically, De Merode and Cowlishaw¹⁰ found that daily sales of protected species in urban markets in Congo increased 5-fold in wartime, with significant increases in the sales volume of large, protected species, including elephants, buffalo, and hippopotamuses. During peacetime, the bushmeat trade in northern Congo was controlled by a small group of female wholesalers who were supported by military officers. The military officers provided protection against fines and extortion (to the wholesalers, hunters, and transporters) and supplied the automatic rifles used to hunt large, protected species. This market structure limited the number of hunters and other actors in the commodity chain and, in turn, the volume of illegal offtake in the park "because the control over the trade that was exerted by the military officers and wholesalers depended heavily on their network of patron-client relationships, and these relationships required time to establish and maintain."¹⁰

This formal, if not legally regulated system changed dramatically during the months of conflict. The military officers fled northern Congo ahead of the advancing rebel army, and consequently, the wholesalers lost their source of authority and were unable to exercise their previous control over the trade. Crucially, in the absence of the patron-client networks established by the military officers and wholesalers, there was no restraint on the number of individuals who could extract, transport, and

sell bushmeat. Lower ranking soldiers (who did not retreat with their officers) began to exploit the potential profits of the bushmeat trade. Depending on the specific job, the number of players involved increased between 2-fold and 6-fold.¹⁰

HUNTING FOR CONSENSUS: SOLUTIONS TO THE BUSHMEAT CRISIS¹¹

Finding a consensus solution to the bushmeat crisis is a difficult task because of the many competing interests involved. Animal rights activists are concerned

Finding a consensus solution to the bushmeat crisis is a difficult task because of the many competing interests involved.

with the moral implications of eating wild animals, particularly species phylogenically close to humans, such as apes. Conservationists are concerned with the preservation of wild species and forests. Advocates of development are concerned with the livelihoods of individuals and the economic growth of countries. In some cases, the harvesting and consumption of bushmeat is a valuable resource for providing inexpensive protein and economic opportunity. Anthropologists are concerned with preserving local traditions. Public health officials are concerned with both limiting zoonotic disease emergence and maintaining food security for rural communities, especially in areas in which livestock production is not viable. How do we reconcile these seemingly disparate interests?

For better or worse, hunting and consuming bushmeat have long been a part of human history, and it seems unlikely that tastes and habits will change dramatically in the near future. Thus, stopping the bushmeat trade on moral grounds is doubtful. Both conservationists and development planners agree that among selected species, harvesting bushmeat at levels that are sustainable would be ideal.¹¹ For most large-bodied species with low reproductive rates, a sustainable bushmeat trade is not feasible; the low level of harvest rates that would be ecologically sustainable would be economically unsustainable.

A sustainable bushmeat trade of selected species, which is both ecologically and economically feasible, would still continue to bring humans and wildlife in close contact, maintaining the risk of zoonotic

disease emergence through increased opportunities for transmission. To address this dilemma, the ideal solution seems to be a properly regulated and sustainable bushmeat trade that is actively monitored by multidisciplinary teams of health experts in order to detect and predict zoonotic disease emergence. So how do we create such a system?

Law Enforcement

Some suggest that the immediate priority for creating a sustainable bushmeat trade is a massive investment in law enforcement.⁷ Fines and enforcement, when properly implemented, have been shown to work in some situations. Researchers in the Serengeti National Park demonstrated a strong correlation between the level of funding for

Fines and enforcement, when properly implemented, have been shown to work in some situations.

antipoaching and the trends in abundance of African buffalo, elephants, and black rhinoceroses.¹² Currently, the legislated fine for hunting protected species in Ghana is 10,000 cedis (1.14 US dollars). Bioeconomic simulations suggests that fines of at least 100,000 cedis, coupled with targeted law enforcement (based on the most vulnerable species), would likely have the most wide-ranging effects.¹³ However, the political will to create strong wildlife management regulations, the ability to establish proper penalties for violations, the infrastructure to enforce laws, and the motivation to enforce laws vary widely among and within countries. In many cases, helping local communities to understand good hunting management practices can accomplish far more than top-down regulation.

Agricultural and Other Development

Because bushmeat is a source of both income and food security for many individuals, it is tempting to suggest that agricultural development would reduce the bushmeat trade. However, economists have found that increases in agricultural returns are ambiguous in their effects on bushmeat hunting. On the one hand, they increase the proportion of labor devoted to agriculture rather than hunting. On the other hand, they can increase the consumption of bushmeat because incomes rise and hunters gain access to more expensive and efficient technology, such as modern weapons to hunt for bushmeat.¹⁴ Substitution of domestic protein is further complicated by environmental realities. In many tropical

areas of the world, traditional livestock production is difficult. Domestic livestock species have been genetically selected for grassland production systems. These species fair poorly in forests. A common alternative is raising cattle on grasslands and then moving them to communities in forested areas as a food source. When they are brought into these areas as a food source, the animals are herded long distances to the destination on foot and then butchered locally as needed because of a lack of vehicle transport and refrigeration for storage of meat. This produces the dilemma of introducing infectious diseases of domesticated animals to naïve populations of wildlife and people in remote areas. Hence, the harvesting of wildlife for human consumption at local community scales provides a suite of advantages but will require appropriate interventions to ensure both the sustainability of the resources and safety from the transmission of disease.

More investment in Africa, in areas that ensure education and political stability and move beyond resource depletion, would go a long way in controlling the bushmeat trade and reducing emerging disease risk. As demonstrated earlier, civil wars

More investment in Africa, in areas that ensure education and political stability and move beyond resource depletion, would go a long way in controlling the bushmeat trade and reducing emerging disease risk.

and logging are 2 of the biggest reasons for the dramatic increase in the bushmeat trade in Africa. Development money from the West plays an important yet complicated role. For example, many of the truck roads and logging projects in Africa are funded by donor agencies or companies based in Europe, Asia, or the Americas.⁴ Outside funding agencies could compel logging and oil companies to control illegal bushmeat hunting, transport, and consumption on their concessions. Funding agencies could withhold project funds from uncooperative companies. It has been suggested that this also could remove some of the burden from national conservation programs.⁷ As an example, the International Finance Corporation of the World Bank Group currently includes requirements and provisions for animal welfare in their agricultural contracts to provide development funds to private companies.¹⁴ By working with private businesses, the International Finance Corporation may make these

requirements without violating any World Trade Organization agreements.¹⁴

Surveillance

In addition to a controlled, sustainable bushmeat trade, some researchers have emphasized the need for surveillance programs, monitored by microbiologists and human and veterinarian health officials, in order to help detect, predict, and prevent zoonotic disease emergence.⁴ One idea is to take paired blood samples from wild animals and their hunters, create

In addition to a controlled, sustainable bushmeat trade, some researchers have emphasized the need for surveillance programs, monitored by microbiologists and human and veterinarian health officials, in order to help detect, predict, and prevent zoonotic disease emergence.

genomic maps of the zoonotic pool, and follow genomic change in zoonotic agents over time.¹⁵ By creating these maps, researchers hope to be able to identify the mutations that allow pathogens to cross species and the mutations that support human-to-human transmission. Ideally, this research would also help public health officials target hotspots of what Wolfe *et al.*⁴ termed *viral chatter*. That is the concept that zoonotic viruses emerge and become established in human populations after a process of repeated transmissions that may not be successful. For example, sequence data from HIV-1 and HIV-2 suggest that as many as 10 prior transmission events into human population occurred over the last century before the virus emerged globally. However, as researchers develop the ability to detect geographic areas of repeated transmission, what do local people do with this information? Should they quarantine these areas? Should they prohibit further hunting?

A response that includes the engagement of local stakeholders as well as the public health community and wildlife authorities is needed if we are to be successful at reducing the risk of emergence of zoonotic diseases, maintaining viable wildlife populations, and supporting the livelihoods of people living in areas rich in biodiversity. Meanwhile, it is clear that contact between humans and animals is changing dramatically both in nature and in volume, and an interdisciplinary approach, combining human and

wildlife biology, microbiology, virology, epidemiology, and anthropology, will be necessary to address the emergence of zoonotic diseases.

A response that includes the engagement of local stakeholders as well as the public health community and wildlife authorities is needed if we are to be successful at reducing the risk of emergence of zoonotic diseases, maintaining viable wildlife populations, and supporting the livelihoods of people living in areas rich in biodiversity.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. World Health Organization. *The Global Burden of Disease: 2004 Update*. Geneva, Switzerland: WHO Press; 2008.
2. Lederberg J, Shope RE, Oakes SC Jr, eds. *Emerging Infections: Microbial Threats to Health in the United States*. Washington, DC: National Academy Press; 1992.
3. Taylor LH, Latham SM, Woolhouse MEJ. Risk factors for human disease emergence. *Philos Trans Biol Sci* 2001; 356: 983–989.
4. Wolfe ND, Daszak P, Kilpatrick AM, Burke DS. Bushmeat hunting, deforestation, and prediction of zoonoses emergence. *Emerg Infect Dis* 2005; 11: 1822–1827.
5. Wolfe ND, Switzer WK, Carr JK, et al. Naturally acquired simian retrovirus infections in central African hunters. *Lancet* 2004; 363: 932–937.
6. Karesh WB, Cook RA, Bennett EL, Newcomb J. Wildlife trade and global disease emergence. *Emerg Infect Dis* 2005; 11: 1000–1002.
7. Walsh PD, Abernethy KA, Bermejo, et al. Catastrophic ape decline in western equatorial Africa. *Nature* 2003; 422: 611–614.
8. Quammen D. Almost cannibalism. *New York Times* June 15, 2003.
9. Auzel P, Wilkie DS. Wildlife use in northern Congo: hunting in a commercial logging concession. In: Robinson JG, Bennett EL, eds. *Hunting for Sustainability in Tropical Forests*. New York, NY: Columbia University Press; 2000; 413–426.
10. De Merode E, Cowlshaw G. Species protection, the changing informal economy, and the politics of access to the bushmeat trade in the Democratic

- Republic of Congo. *Conserv Biol* 2006; 20: 1262–1271.
11. Bennett EL, Blencowe E, Brandon K, et al. Hunting for consensus: reconciling bushmeat harvest, conservation, and development policy in West and Central Africa. *Conserv Biol* 2007; 21: 884–887.
 12. Hilborn R, Arcese P, Borner M, et al. Effective enforcement in a conservation area. *Science* 2006; 314: 1266.
 13. Damania R, Milner-Gulland E, Crookes D. A bioeconomic analysis of bushmeat hunting. *Proc R Soc Lond B Biol Sci* 2005; 272: 259–266.
 14. *Good Practice Note: Animal Welfare in Livestock Operations*. Washington, DC: International Finance Corp.; 2006.
 15. Brilliant L, Birmingham C, Canadien V, et al. Q&A: Larry Brilliant. Interviewed by Declan Butler. *Nature* 2008; 451: 379.

Xenotransplantation, Xenogeneic Infections, Biotechnology, and Public Health

Louisa E. Chapman, MD, MSPH

Office of Critical Information Integration and Exchange, Centers for Disease Control and Prevention,
Atlanta, GA

OUTLINE

WHY IS XENOTRANSPLANTATION A PUBLIC HEALTH CONCERN?

Rationale for xenotransplantation

Why are public health professionals interested in xenotransplantation?

XENOTRANSPLANTATION AND PUBLIC HEALTH: 1995

Public Health Guidance

Endogenous Retroviruses

2009: WHAT IS THE FUTURE OF XENOTRANSPLANTATION?

to infect human cells, and have generated public health concern. The history of xenotransplantation, the implications for public health, the global consensus on public safeguards necessary to accompany clinical trials, and the future direction of xenotransplantation are discussed in the context of public health. *Mt Sinai J Med* 76:435–441, 2009. © 2009 Mount Sinai School of Medicine

Key Words: porcine endogenous virus, public health, xenogeneic infections, xenotransplantation.

ABSTRACT

Xenotransplantation is the attempt to use living biological material from nonhuman animal species in humans for therapeutic purposes. Clinical trials and preclinical studies have suggested that living cells and tissue from other species have the potential to be used in humans to ameliorate disease. However, the potential for successful xenotransplantation to cure human disease is coupled with the risk that therapeutic use of living nonhuman cells in humans may also serve to introduce xenogeneic infections of unpredictable significance. Animal husbandry practices and xenotransplantation product preparation may eliminate most exogenous infectious agents prior to transplantation. However, endogenous retroviruses are present in the genomes of all mammalian cells, have an inadequately defined ability

Xenotransplantation can be briefly described as an attempt to use living biological material from nonhuman animals in humans for therapeutic benefit. The US Public Health Service defines xenotransplantation more formally as the transplantation, implantation, or infusion into a human recipient of either (1) live cells, tissues, or organs from a nonhuman animal source or (2) human body fluids, cells, tissues, or organs that have had ex vivo contact with live nonhuman animal cells, tissues, or organs.¹

Xenotransplantation may sound like an unlikely event. However, in vitro fertilization, a practice by which infertile couples are often able to bear children via the removal of eggs and sperm from the intended parents, fertilization of the eggs in a laboratory, growth of those fertilized eggs to a multicell stage over 3 to 5 days, and implantation of the eggs into the mother's uterus, is no longer an unlikely event. In the 1990s, the substrate used in the laboratory to support the development of a fertilized egg into the multicell stage was frequently a cell line of nonhuman origin. Thus, the multicell stage fertilization product implanted into the mother's uterus and ultimately the resulting infant was, by US Public Health Service definition, a xenotransplantation product. Additionally, hundreds of patients have been treated with investigational xenotransplantation products intended (1) to sustain patients suffering hepatic failure until a liver is available for transplant (hemoperfusion through

Address Correspondence to:

Louisa E. Chapman

Office of Critical Information
Integration and Exchange
Centers for Disease Control and
Prevention Atlanta, GA
Email: lec3@cdc.gov

a porcine liver or hepatocytes), (2) to decrease the dependence of diabetics on insulin (porcine pancreatic cell implants), or (3) to improve functions in patients with Parkinson's disease (implantation of porcine neurological cells) or other functions.^{2,3}

WHY IS XENOTRANSPLANTATION A PUBLIC HEALTH CONCERN?

Rationale for Xenotransplantation

As transplantation surgery has become more technically proficient, the factor limiting transplant patient survival has ceased to be the technical complexity of the surgery and has instead become the limited availability of donor organs appropriate for transplantation. According to the United Network for Organ Sharing Web site, on June 3, 2009 at 4:45 PM, 102,010 people were candidates for the organ transplantation waiting list. However, between January 1 and May 29, 2009, only 2304 organ donors had become available. Looking specifically at heart transplantation, we find that between July 1, 2007 and June 30, 2008, only 12% of heart transplant recipients failed to survive the first year after surgery. During that same period, 18% of the people who were both in need of a heart transplant and on the waiting list died while awaiting transplantation. Thus, as a result of the mismatch between the demand for and supply of donor organs for transplantation, the greatest risk of dying due to heart transplantation today is attributable to the scarcity of transplantable organs rather than to the risk of either surgery or posttransplantation organ rejection.⁴

The shortage of donor organ availability first drove interest in exploring whether organs from nonhuman animals could be adapted for transplantation into humans. Xenotransplantation has been envisioned as a source of spare-part organs to be transplanted as a definitive cure for end organ failure. However, xenotransplantation has also been envisioned as a bridge therapy: an interim therapeutic step intended only to sustain the patient long enough to allow a compatible human donor organ to become available.

The disparity between the demand for transplantation generated by organ failure and the supply of donated human organs was the original driving force behind efforts to develop xenotransplantation. However, a second driving force emerged after it was recognized that differences in species susceptibility to specific infections might be exploited to human advantage. Two early landmark experiments demonstrated the application of this concept.

When Dr. Thomas Starzl, a pioneering transplant surgeon in Pittsburgh, learned that baboons were refractory to infection with either hepatitis B or human immunodeficiency virus (HIV), he wondered if this species difference in susceptibility could be exploited to the advantage of HIV-infected patients who were dying of liver failure due to hepatitis B virus infection. In 1992, following advances in the ability to control both cellular and humoral components of xenograft rejection in vitro, Dr. Starzl and his colleagues attempted to transplant a baboon liver into a 35-year-old HIV-infected man with hepatitis B virus-associated chronic active hepatitis. The patient survived with little evidence of rejection, and products of hepatic synthesis became those of the baboon liver without evidence of an obvious adverse impact. The patient died on day 70 after transplantation because of a cerebral and subarachnoid hemorrhage caused by an angioinvasive *Aspergillus* infection. Although Dr. Starzl did not repeat this experiment, he concluded that this experiment had demonstrated the feasibility of controlling the rejection of baboon livers transplanted into human recipients.⁵ Conceptually, he had ushered in a new era of thought about xenotransplantation.

On December 14, 1995, a 38-year-old activist who had been infected with HIV for more than 15 years underwent a controversial experiment. His bone marrow was suppressed (not ablated) by sublethal doses of radiation and chemotherapeutic drugs, after which he received an infusion of stem cells and facilitator cells procured from the bone marrow of a baboon. Facilitator cells, discovered by Dr. Suzanne Ildstad, another Pittsburgh surgeon, appear to allow stem cells to proliferate in other species without producing graft-versus-host disease.⁶ This experimental attempt to reconstitute chimeric functional bone marrow in a patient with acquired immune deficiency syndrome (AIDS) was another conceptualization of how the species differences in susceptibility to infection could be exploited to human health advantage. This experiment occurred prior to the advent of cocktail antiretroviral therapy at a point when the standard treatment approaches were not working adequately and the AIDS activist community was increasingly convinced that radical new approaches were necessary. In this atmosphere of desperation, a US Food and Drug Administration advisory committee debated and then voted to allow this controversial human experiment that might actually foreshorten rather than prolong the patient's life. The experiment went forward, and the patient survived and improved for reasons that are unclear, as the baboon bone marrow cells were not identifiable in his bone marrow beyond the first month.⁷

These 2 experiments illustrate a transition in the conceptual nature of xenotransplantation. The first concept, that whole animal organs were anticipated to serve as spare-part replacements for failed human organs, was driven by the inability of the supply of human organs to keep pace with the ability of technological advances in transplantation to save lives. Increasingly since 1990, the development of xenotransplantation applications has been influenced by the recognition that unique properties of therapeutic materials originating from nonhuman species may be exploited to the advantage of human health. Increasingly, the products are cellular rather than whole organs. Problems with immune rejection and the absence of preclinical data meeting specific recommendations for the survival of xenogeneic organs in nonhuman primates prior to clinical use in humans further influence preferential interest in cellular transplantation versus organ transplantation. Recent success with islet allotransplantation for diabetes is driving renewed interest in using porcine islets because of a lack of a sufficient supply of human islets for allotransplantation. The transplantation of cellular products likely represents the most viable near future of xenotransplantation.

Why Are Public Health Professionals Interested in Xenotransplantation?

Xenotransplantation first came to the compelling attention of public health authorities in the mid-1990s, around the time of the previously described experiments by Dr. Starzl and Dr. Ildstad.⁸ The combination of advances in the control of immune rejection and the engineering of transgenic pigs that contained certain human genes anticipated to improve porcine xenotransplantation product survival in human recipients led to great enthusiasm for the rapid movement of xenotransplantation applications into human clinical trials. The absence of any precedent for regulatory policy for xenotransplantation clinical trials was of pressing concern to the public health authorities charged with safety oversight.

Although Dr. Starzl, Dr. Ildstad, and their colleagues were focused on the potential to transform the health of individuals, the public health focus was on the protection of community health. Some xenotransplantation proposals could provide great societal benefit if successful. For example, in 2005–2006, the crude prevalence of total diabetes in US residents 20 years old or older was 12.9%.⁹ For decades before human insulin became available, diabetes was managed through the intermittent injection of porcine insulin. Imagine the impact if diabetes could be functionally cured through the infusion of functioning

porcine pancreatic islet cells. One modeling study estimated the health economic impact of maintaining glycosylated hemoglobin values in all US patients with currently uncontrolled type 1 or type 2 diabetes mellitus at the American Diabetes Association standard of 7.0% and at the American Association of Clinical Endocrinologists target of 6.5%. This analysis, run from a societal perspective over a 10-year time horizon, estimated that achieving this level of maintenance could achieve total direct medical cost savings of 35 to 50 billion US dollars, respectively, over 10 years. When indirect cost savings were included, the total savings increased to 50 to 72 billion US dollars, which corresponded to 4% to 6% of the total annual US health care costs of 1.3 trillion US dollars.¹⁰ This analysis estimated only financial savings and did not address quality of life issues or years of productive life reclaimed.

Although xenotransplantation's potential for positive benefits to individuals and society was a primary driver for the clinical research community, the attention of the public health community was captured more by concern for the potential for unintended negative consequences. Xenotransplantation is intended to benefit the health of individuals by replacing nonfunctioning or malfunctioning human cells, tissues, or organs with functioning nonhuman animals cells, tissue, or organs. However, the implantation, transplantation, or infusion of living nonhuman tissue into humans for therapeutic purposes has an associated potential to also transfer infections across species lines into humans. Because xenotransplantation applications breach normal host defenses and are frequently accompanied by pharmacological immune suppression, xenotransplantation may be ideally suited to have the unintended consequence of introducing new infections into the human recipients. The potential for implanted living nonhuman animal cells to also transfer infections across species lines into the human population (xenozoonosis or xenogeneic infections) was a major concern. Zoonotic infections occur in nature and produce disease in individuals and epidemics in human populations. Examples of zoonotic epidemics include the 1993 hantavirus pulmonary syndrome outbreak in the American southwest¹¹ and the epidemic of encephalitis that followed the introduction of West Nile virus into New York in 1999.¹² Zoonotic infection of humans by avian influenza viruses caused the largest recorded pandemic in human history, the Influenza Pandemic of 1918.¹³ AIDS, now understood to have resulted from the introduction of a simian immunodeficiency virus infection across species lines into humans,¹⁴ has gone through all these stages and is no longer a zoonosis, an epidemic, or a pandemic.

AIDS now is simply an endemic infection affecting all human populations throughout the world, and as much as any single force in this century, it is reshaping the world.

This then is the primary basis for public health interest in xenotransplantation and related biotechnologies. The potential for xenotransplantation to benefit individual patients is inevitably linked to a potential to introduce harm to human populations through the unintended introduction of xenogeneic infections. Because infectious diseases do not respect geopolitical boundaries, a xenotransplantation clinical trial anywhere that is not accompanied by adequate public safeguards is a concern to the global community.

XENOTRANSPLANTATION AND PUBLIC HEALTH IN 1995

In 1995, at the request of Phil Lee, the Assistant Secretary for Health of the Department of Health and Human Services, federal agencies first began to examine xenotransplantation as a public health issue. At that time, the potential for new biotechnical approaches to alleviate human suffering by the use of living biological material from nonhuman animals was recognized. These experimental approaches had the exciting potential to have an unprecedented positive impact on a broad spectrum of human disease. However, these approaches also carry an unquantifiable and probably small but still existent risk of unintended negative side effects via the introduction of xenogeneic infections into the human population.

Recognition of the potential for xenotransplantation to have a negative public health impact due to xenogeneic infections inspired collaborative efforts by the academic, clinical, industrial, research, and public health communities to identify ways to develop this promising biotechnology while adequately safeguarding the health of the larger community. The first public health priority was the development of an international consensus on the public safeguards necessary to allow xenotransplantation clinical trials to proceed with public confidence. Once consensus was achieved, it was implemented through the development of appropriate public health policy translated into regulatory practice. Simultaneous research efforts explored clinical interventions and bioengineering approaches that might increase the safety of clinical trials. The stakeholder communities also undertook collaborative basic science research to better define fundamental understandings of both the risk and potential of xenotransplantation.

Public Health Guidance

Public health guidance documents are available from national health authorities in most nations in which clinical xenotransplantation trials have been undertaken or considered, including the United States, Canada, multiple European countries, the European Union, and Australia, and from multinational organizations such as the World Health Organization and the Organization for Economic Cooperation and Development.^{1,15} These various guidances are interpretations of a single global consensus.

The US Public Health Service and other guidances build on a foundation that requires xenotransplantation product source animals to originate from closed colonies of purpose-bred animals with husbandry practices that limit and define the lifelong exposures of these animals. These guidances emphasize the importance of pretransplantation screening of source animals and herds to identify and eliminate problematic infectious agents prior to the development of xenotransplantation products and of post-transplantation surveillance of xenotransplantation recipients to identify and contain any xenogeneic infections. Posttransplantation surveillance is necessary to identify infectious agents that (1) were transplanted because they were not known to exist at the time of transplant screening (eg, severe acute respiratory syndrome-associated coronavirus prior to 2003), (2) were known to exist but could not be detected because of inadequate diagnostic tools (eg, prions), or (3) were known to be present in the source animal and could not be removed (eg, endogenous retroviruses).

Endogenous Retroviruses

Retroviruses are RNA viruses that replicate by transcribing viral RNA into DNA by reverse transcription.¹⁶ Proviral DNA is integrated into the host cell genome and replicated with the host cell DNA. Exogenous retroviruses exist as independent cell invaders that are transmitted horizontally by infection. The most familiar example is HIV.

However, when retroviruses become integrated into the host cell genome within a germ cell, they can then be transmitted vertically by inheritance through the germline DNA. These endogenous retroviruses exist as proviral DNA integrated into the germlines of all mammals, including the genomes of humans and all species considered as source animals for xenotransplantation. Endogenous retroviruses represent a sort of fossil remnant of what are presumed to have once been exogenous retroviruses that integrated into the host germline eons ago and remain as an inherited part of the genetic structure of every cell

of the species today. These endogenous retroviruses may express infectious viruses but do not cause disease in the host species. However, many endogenous retroviruses are xenotropic; this means that they are able to infect cells from other species. Endogenous retroviruses of pigs and baboons are able to infect human cells in vitro. Thus, living animal tissue that is apparently devoid of exogenous infectious agents nonetheless retains an innate infectious potential due to the presence of endogenous retroviruses.

C-type particles (crescent-shaped formations on the membranes of cells associated with the budding of C-type retroviruses) expressed from a variety of porcine cell lines were identified during the 1970s and 1980s and characterized as endogenous retroviruses capable of infecting ST-Iowa cells (a cell line derived from *Sus scrofa*) in vitro. After 1995, concerns about endogenous retroviruses in xenotransplantation products inspired studies that defined infectivity and host ranges for porcine endogenous retrovirus (PERV).^{17,18} PERV is expressed from multiple porcine cell lines and primary tissues. Of 3 identified variants, 2 (PERV-A and PERV-B) productively infect multiple human cell lines, although human peripheral blood mononuclear cells appear resistant to productive infection. Both the murine leukemia virus and feline leukemia virus share more than 60% homology with PERV. This homology has been exploited to develop serologic assays for PERV, and the recognized characteristics of these viruses have been used as a basis for reasoning by analogy about how PERV may behave biologically.^{2,3,13,18}

In response to these findings, on October 16, 1997, the US Food and Drug Administration placed all xenotransplantation trials using porcine products in the United States on clinical hold. Reimplementation of clinical trials required the development of assays for detection of infectious PERV in xenotransplantation products, implementation of surveillance for PERV infections in recipients, and development of informed consent documents adequately informing clinical trial participants of potential risks associated with the presence of PERV in porcine xenotransplantation products.¹⁹ In 1999, in recognition of a global public health consensus, the US Food and Drug Administration issued additional guidances that preclude the use of nonhuman primates as source animals for xenotransplantation products and that defer xenotransplantation recipients from the donation of blood and other biological materials as precautionary measures.²⁰

New tools were necessary to enable laboratory surveillance for endogenous retrovirus infection. Because PERV DNA is a normal part of the genome of every porcine cell, polymerase chain reaction

(PCR) identification of PERV DNA will be inevitable whenever transplanted porcine cells are present. The ability to discriminate PERV infection from the presence of PERV DNA-containing porcine cells in a xenotransplantation product recipient requires additional testing. Most approaches combine a PCR assay for PERV DNA with a PCR assay for a marker of porcine (or other source animal) DNA. Although these assays frequently identify source animal mitochondrial DNA, other repetitive sequences such as centromeric sequences have also been used.^{21–25} The presence of PERV DNA in the absence of porcine mitochondrial DNA implies PERV infection, whereas in the presence of porcine mitochondrial DNA, it simply implies that not all xenogeneic cells have been rejected by the recipient.^{21–25} This basic approach is further refined with quantitative tests that assess the relative abundance of PERV DNA and host mitochondrial DNA in biological material from recipients with respect to the ratio that existed in source animal cells prior to transplantation. These, combined with reverse-transcriptase PCR tests that identify PERV-specific RNA evidence of viral expression, Amp-RT or similar assays that identify generic reverse-transcriptase activity, and western blot assays that identify seroreactivity against homologous retroviruses, compose the basic armamentarium for laboratory surveillance for PERV infection in porcine xenotransplant recipients.^{21–25}

An early series of retrospective studies failed to identify evidence of PERV infection in recipients of porcine xenotransplantation products. Two landmark negative studies were published in *Lancet* in 1998. Patience *et al.*²³ found no evidence of PERV infection in 2 patients who had experienced short-term extracorporeal perfusion of their blood through pig kidneys in the absence of immunosuppression. Heneine *et al.*²⁴ did not identify PERV infection in 10 immunosuppressed diabetic recipients of human kidney transplants who also received fetal porcine islet cells either infused into the portal vein or inserted under the capsule of the kidney allograft; porcine cells had persisted for up to 6 months following transplantation. In 1999, the negative results of a global collaboration by the xenotransplantation research community were published.²⁵ This collaboration studied a large number of patients ($n = 160$) who were exposed to porcine xenotransplantation products through a wide variety of methods. PERV infection was sought through double-blinded testing in 2 laboratories using independently developed assays. Microchimeric porcine cells were identified in the peripheral blood of 23 recipients up to 8.5 years after exposure ended. This finding, while

unexpected, is consistent with evidence of Y chromosome-containing microchimeric cells in the peripheral blood of women decades after they gave birth to male offspring.²⁰ In the context of xenotransplantation, this finding signifies that transient exposure to xenotransplantation products may result in persistent exposure to risk of PERV infection.

2009: WHAT IS THE FUTURE OF XENOTRANSPLANTATION?

Since the dialogue between public health and xenotransplantation began in the mid-1990s, much effort has been applied to studies attempting to elucidate aspects of xenotransplantation relevant to infectious risk.²⁷ The development of a global agreement on public safeguards that should accompany xenotransplantation clinical trials was a significant advance. A growing body of evidence has failed to identify PERV infection in xenotransplantation product recipients, although recent advances intended to overcome immune rejection (ie, the development of alpha-gal knockout source pigs) may inadvertently increase the risk of PERV transmission.²⁸ Modifications of xenotransplantation product bioengineering have been shown to diminish the release of PERV virions in vitro; this observation suggests that such modifications may reduce a recipient's risk of exposure to infectious PERV in vivo.²⁹ Other in vitro experiments suggest that transient use of antiretrovirals peri-transplant may increase selective pressure against persistent infection.³⁰ The recognized homology between PERV and feline leukemia virus suggests that effective vaccination may be possible, although other observations suggest that vaccines which protect against PERV infection may also contribute to the rejection of porcine xenotransplantation products.³¹ Although much remains to be explored in all these areas, findings to date endorse the decision by regulatory authorities to allow xenotransplantation clinical trials to proceed with public safeguards in place and diminish, but do not eliminate, concern about the unintended introduction of new infections into the human population as a byproduct of efforts to cure individual disease.

The potential for xenotransplantation to introduce xeno-genetic infections remains a concern but is no longer the rate-limiting step in advancement of the field of xenotransplantation. Whether xenotransplantation will deliver on the promise raised by early visions will depend more on whether basic research can overcome the remaining immune barriers to long-term survival of xenotransplantation products in vivo, on discoveries about the adequacy of xenoproduct

physiological function in humans, and on whether advances in the related fields of cloning and regenerative medicine outpace the rate of discovery in xenotransplantation to the point of irrelevancy.

ACKNOWLEDGMENT

This article is an adaptation of a lecture given by the author at Mount Sinai School of Medicine for the Master of Public Health program "Special Topics in Zoonosis" on April 10, 2008.

The findings and conclusions in this presentation represent the views of the author alone and not those of the Centers for Disease Control and Prevention, the US Public Health Service, the US Government, or any other entity.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. United States Public Health Service. Public health service guideline on infectious disease issues in xenotransplantation. *MMWR Recomm Rep* 2001; 50: 1–46.
2. Chapman LE, Bloom ET. Clinical xenotransplantation. *JAMA* 2001; 285: 2304–2306.
3. Boneva, R., Folks TM, Chapman LE. Infectious disease issues associated with xenotransplantation. *Clin Microbiol Rev* 2001; 14: 1–14.
4. United Network for Organ Sharing. Available at: <http://www.unos.org>. Accessed June 2009.
5. Starzl TE, Fung J, Tzakis A, et al. Baboon to human transplantation. *Lancet* 1993; 341: 65–71.
6. Gaines BA, Colson YL, Kaufman CL, Ildstad ST. Facilitating cells enable engraftment of purified fetal liver stem cells in allogeneic recipients. *Exp Hematol* 1996; 24: 902–913.
7. Ildstad ST. Xenotransplantation for AIDS. *Lancet* 1996; 347: 761.
8. Chapman LE, Salomon DR, Patterson AP, et al. Xenotransplantation and xeno-genetic infections. *N Engl J Med* 1995; 333: 1498–1501.
9. Cowie CC, Rust KF, Ford ES, et al. Full accounting of diabetes and pre-diabetes in the U.S. population in 1988–1994 and 2005–2006. *Diabetes Care* 2009; 32: 287–294.
10. Minshall ME, Roze S, Palmer AJ, et al. Treating diabetes to accepted standards of care: a 10-year projection of these estimated economic and health impact in patients with type 1 and type 2 diabetes mellitus in the United States. *Clin Ther* 2005; 27: 940–950.
11. Chapman LE, Khabbaz RF. Review: etiology and epidemiology of the Four Corners hantavirus outbreak. *Infect Agents Dis* 1994; 3: 234–244.

12. Pealer LN, Marfin AA, Petersen LR, et al. Transmission of West Nile virus through blood transfusion in the United States in 2002. *N Engl J Med* 2003; 349: 1236–1245.
13. Vana G, Westover KM. Origin of the 1918 Spanish influenza virus: a comparative genomic analysis. *Mol Phylogenet Evol* 2008; 47: 1100–1110.
14. Hahn BH, Shaw GM, De Cock KM, et al. AIDS as a zoonosis: scientific and public health implications. *Science* 2000; 287: 607–614.
15. *Xenotransplantation: Guidance on Infectious Disease Prevention and Management*. Geneva, Switzerland: World Health Organization; 1998.; WHO document WHO/EMC/ZOO/98.1.
16. Mahy BWJ. *The Dictionary of Virology*. 4th ed. London, United Kingdom: Academic Press; 2009; 407.
17. Brown J, Matthews L, Sandstrom PA, Chapman LE. Xenotransplantation and the risk of retroviral zoonosis. *Trends Microbiol* 1998; 6: 411–415.
18. Chapman LE, Switzer WM, Sandstrom PA, et al. Infectious disease issues in xenotransplantation. In: Scheld WM, Craig WA, Hughes JM, eds. *Emerging Infections 3*. Washington, DC: ASM Press; 1999.
19. Bloom, ET. Xenotransplantation—Federal Regulatory Considerations. In: Salomon DR, Wilson C, eds. *Current Topics in Microbiology: Xenotransplantation*. Berlin: Springer-Verlag; 2003; 239–251.
20. US Food and Drug Administration. Draft guidance for industry: precautionary measures to reduce the possible risk of transmission of zoonoses by blood and blood products from xenotransplantation product recipients and their contacts. Available at: <http://www.fda.gov/biologicsbloodvaccines/guidancecomplianceregulatoryinformation/guidances/blood/ucm075125.htm>. Accessed June 2009.
21. Heneine W. Strategies for diagnosis of xenotransplant-associated retroviral infections. *Mol Diagn* 1996; 1: 255–260.
22. Heneine W, Switzer WM. Highly sensitive and specific polymerase chain reaction assays for detection of baboon and pig cells following xenotransplantation in humans. *Transplantation* 1996; 62: 1360–1362.
23. Patience C, Patton GS, Takeuchi Y, et al. No evidence of pig DNA or retroviral infection in patients with short-term extracorporeal connection to pig kidneys. *Lancet* 1998; 352: 699–701.
24. Heneine W, Tibell A, Switzer WM, et al. No evidence of infection with the porcine endogenous retrovirus in human recipients of porcine islet cell xenografts. *Lancet* 1998; 352: 695–699.
25. Paradis K, Langford G, Long Z, et al. Search for cross species transmission of porcine endogenous retrovirus in patients treated with living pig tissue. *Science* 1999; 285: 1236–1241.
26. Bianchi DW, Zickwolf GK, Weil GJ, Sylvester S. Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum. *Proc Natl Acad Sci* 1996; 93: 705–708.
27. Chapman LE. Xenotransplantation. In: Burroughs T, Knobler S, Lederberg J, eds. *The Emergence of Zoonotic Diseases: Understanding the Impact on Animal and Human Health*. Washington, DC: National Academy Press; 2001; 17–20.
28. Chapman LE, Wilson CA. Commentary: implications of the advent of homozygous α 1,3-galactosyltransferase gene-deficient pigs on transmission of infectious agents. *Xenotransplantation* 2003; 10: 287–288.
29. Nyberg SL, Hibbs JJR, Hardin JA, et al. Transfer of porcine endogenous retrovirus across hollow fiber membranes: significance to a bio-artificial liver. *Transplantation* 1999; 67: 1251–1255.
30. Quri SH, Magre S, Garcia-Lerma JG, et al. Susceptibility of the porcine endogenous retroviruses to reverse transcriptase and protease inhibitors. *J Virol* 2001; 75: 1048–1053.
31. Simeonovic CJ, Ziolkowski AF, Popp SK, et al. Porcine endogenous retrovirus encodes xenoantigens involved in porcine cellular xenograft rejection by mice. *Transplantation* 2005; 79: 1674–1682.

Outbreak of Progressive Inflammatory Neuropathy Following Exposure to Aerosolized Porcine Neural Tissue

Tracy M. DeAngelis, MD, and Liang Shen, BS

Mount Sinai School of Medicine, New York, NY

OUTLINE

EARLY INVESTIGATION

Focus on Chronic Inflammatory Demyelinating Polyneuropathy

LATER INVESTIGATIONS

Focus on Progressive Inflammatory Neuropathy

FOLLOW-UP AND CURRENT STATUS

ABSTRACT

In the fall of 2007, the Minnesota Department of Health was notified of 11 cases of an unexplained neurological illness, all linked to a pork processing plant, Quality Pork Processors, Inc., in Austin, MN. The cluster of workers had been experiencing similar symptoms, including fatigue, pain, numbness, and tingling in their extremities as well as weakness. The symptoms were described as more sensory than motor, and all patients had evidence of polyradiculoneuropathy with signs of nerve root irritation. An epidemiological investigation revealed that the only commonality between cases was their exposure to a pork brain extraction procedure involving compressed air. As relatives of the cases remained asymptomatic and all cultures for known pathogens were negative, the etiology of the syndrome seemed not to be infectious. Clinically, the syndrome was most akin to chronic inflammatory demyelinating polyneuropathy. Laboratory tests corroborated the clinical findings, revealing inflammation of peripheral nerves and nerve roots; however, these cases also had features

clinically distinct from chronic inflammatory demyelinating polyneuropathy as well as laboratory testing revealing a novel immunoglobulin G immunostaining pattern. This suggested that the observed inflammation was the result of 1 or more unidentified antigens. This syndrome was ultimately dubbed *progressive inflammatory neuropathy* and was theorized to be an autoimmune reaction to aerosolized porcine neural tissue. Since the investigation's outset, 18 cases of progressive inflammatory neuropathy have been identified at the Minnesota pork processing plant, with 5 similar cases at an Indiana plant and 1 case at a Nebraskan plant. The plants in which cases have been identified have since stopped the use of compressed air in removing pork brains. All cases have stabilized or improved, with some requiring immunosuppressive and analgesic treatment. The study of progressive inflammatory neuropathy is ongoing, and the details of this investigation highlight the value of epidemiological principles in the identification and containment of outbreaks while researchers attempt to uncover the unique pathophysiology and potential etiology of the illness. *Mt Sinai J Med* 76:442–447, 2009. © 2009 Mount Sinai School of Medicine

Key Words: autoimmune, chronic inflammatory demyelinating polyneuropathy, neuropathy, polyradiculoneuropathy, pork plant workers, progressive inflammatory neuropathy, sensorimotor peripheral neuropathy.

In the fall of 2007, the Minnesota Department of Health was notified of 11 cases of an unexplained neurological illness, all linked to a pork processing plant, Quality Pork Processors, Inc., in Austin, MN.¹ A cluster of workers had been experiencing symptoms since late 2006, which included fatigue, pain, numbness, and tingling in their extremities as well as weakness.² There was no evidence of an antecedent illness. These symptoms then progressed and persisted to the time of the Minnesota Department of Health investigation of the outbreak.¹

Address Correspondence to:

Liang Shen, BS

Mount Sinai School of Medicine
New York, NY

Email: liang.shen@mssm.edu

Together with the Centers for Disease Control and Prevention and the Mayo Clinic in Rochester, MN, state public health officials launched an ongoing investigation. To investigators, this constellation of symptoms suggested a possible peripheral nervous system disorder, such as a neuropathy or polyradiculopathy, versus one involving the central nervous system, such as transverse myelitis. Clinical and laboratory investigations would eventually confirm this suspicion, identifying all cases as having evidence of a polyradiculoneuropathy as well as a few cases with cranial nerve and spinal cord involvement. By January 2008, a total of 12 workers had been identified with the aforementioned syndrome. Curiously, all 12 workers reported working either near or at the head table, an area in which swine heads were processed and brains were removed with compressed air.³ Despite this clue, which prompted the plant to voluntarily discontinue the compressed air brain extraction procedure, investigators were no closer to diagnosing the cause of the illness.

EARLY INVESTIGATION

The fact that all cases of this mysterious polyneuropathy occurred in workers at a pork processing plant immediately suggested the possibility of an occupational exposure or disease passed from pigs to humans, such as a zoonotic disease.

The fact that all cases of this mysterious polyneuropathy occurred in workers at a pork processing plant immediately suggested the possibility of an occupational exposure or disease passed from pigs to humans.

A variety of potential infectious pathogens are known to cause peripheral neuropathies and/or transverse myelitides. In fact, infectious neuropathy comprises the largest group of neuropathies in the world.⁴ The causes of infectious neuropathy are vast and include viral, bacterial, and parasitic organisms. For example, human immunodeficiency virus, cytomegalovirus, and hepatitis C are leading viral causes of neuropathy. *Mycobacterium leprae*, the bacterium responsible for leprosy, and *Borrelia burgdorferi*, the vector-borne bacteria that causes Lyme disease, are also major causes of neuropathy.⁴ With the only commonality among cases being exposure to this

pork brain removal procedure, one strong suspect was the *Trichinella spiralis* roundworm, which frequently enters human hosts through ingestion of undercooked pork. Trichinosis, however, classically causes inflammatory myopathy rather than peripheral neuropathy. Another possibility, in light of the uniform exposure to animal brain tissue, was a prion disease or spongiform encephalopathy. Similarly, however, prion diseases are associated clinically not with peripheral neuropathy but rather with precipitous dementia, seizures, encephalopathy, and eventual death; therefore, this possibility was also very unlikely.

In addition to infections directly affecting peripheral nerves, neuropathies that occur as sequelae of a prodromal microbial infection should also be considered. For example, Guillain-Barré syndrome (GBS), a polyneuropathy that frequently follows an antecedent microbial infection, was a strong consideration. GBS is characterized by inflammation and demyelination of peripheral nerves (hence its alternative moniker *acute inflammatory demyelinating polyradiculoneuropathy*).⁵ Some infections that have been identified as preceding GBS include *Campylobacter jejuni*, cytomegalovirus, *Mycoplasma pneumoniae*, Epstein-Barr virus, and influenza virus.⁵ GBS itself has not been known to spread to humans directly from infected pig tissue. Clinically, GBS generally progresses quickly, presenting with acute ascending motor weakness and a loss of deep-tendon reflexes, and is frequently preceded by sensory paresthesias. Patients can also experience dysautonomias such as cardiac dysrhythmias and urinary retention. More than 90% of patients reach the clinical nadir within 4 weeks of symptom onset.⁵ In contrast, the plant workers' symptoms appeared on average to have outlasted this time frame.

Focus on Chronic Inflammatory Demyelinating Polyneuropathy

Because the symptoms persisted for months, a more plausible candidate for the classification of the pork workers' neurological syndrome was chronic inflammatory demyelinating polyneuropathy (CIDP). This possibility had been put forth since the beginning of the investigation.¹ CIDP tends to progress at a slower rate than GBS and often observes a relapsing course.⁶ It can occur at any age, with peak prevalence between 40 and 60 years of age, and demonstrates a slight male predominance.⁶ The symptoms of CIDP are classically defined as weakness in a proximal and distal pattern, usually symmetric, involving both upper and lower extremities, and they present for at least 2 months.⁶

Because the symptoms persisted for months, a more plausible candidate for the classification of the pork workers' neurological syndrome was chronic inflammatory demyelinating polyneuropathy (CIDP).

Facial and neck flexor muscles are also often affected.⁶ In addition to weakness, some patients experience sensory symptoms such as numbness, tingling, and pain.⁶ In contrast to GBS, autonomic function in CIDP is normal.⁷

Although the exact etiology of CIDP has not been established,⁶ a process of autoimmunity has been proposed, involving both cell-mediated and antibody-mediated responses to glycolipid and myelin protein antigens.⁷ Initially, an accumulation of autoreactive CD4+ and CD8+ T cells in the neural endothelium occurs, with macrophage accumulation in the inflammatory fluid. A subsequent release of tumor necrosis factor alpha, interferon gamma, and interleukin 2 results in an increase in endothelial permeability, allowing increased infiltration of antimyelin antibodies. These antibodies, which have been variously described as immunoglobulin M or immunoglobulin G, bind myelin sheaths and induce macrophage phagocytosis of myelin via Fc receptor mechanisms, causing the characteristic pathologic inflammation and demyelination.⁷

Solely using laboratory tests in making a diagnosis of CIDP tends to be difficult because of low sensitivity, and many guidelines today call for using clinical features before laboratory tests.⁶

That said, the laboratory tests used most commonly to aid in the diagnosis of CIDP include serum and cerebrospinal fluid (CSF) testing, electrophysiological testing, and nerve biopsy. Serologic tests can help exclude possible infectious, collagen vascular, neoplastic, and toxic-metabolic mimickers.⁶ CSF studies appear to be the most sensitive test, with over 90% of CIDP patients having elevated CSF protein concentrations.⁶ In about 50% of patients with CIDP, electrophysiological findings demonstrate a typical demyelinating pattern in at least 2 motor nerves reflected by slowed motor conduction velocity, prolonged distal motor latencies, temporal dispersion, delayed or absent F-waves, and/or conduction block.^{6,7} In general, nerve biopsies of patients with an inflammatory demyelinating neuropathy such as CIDP can be difficult to distinguish from those of noninflammatory neuropathies or even normal

controls.⁶ Less than 20% of patients demonstrate both demyelination and axonal changes, and 10% or less show signs of inflammation.⁶ However, biopsies can be used to rule out other disease processes, such as toxic, metabolic, vasculitic, or hereditary neuropathies.⁶

Corticosteroids have been the mainstay of treatment for CIDP since 1958.⁶ Other acceptable substitutes for first-line treatment are intravenous immunoglobulin and plasma exchange.^{6,7} Corticosteroids, such as prednisone, are used in higher doses initially. Once improvement has plateaued or is complete, doses are reduced every few weeks, but the steroids are usually continued indefinitely to prevent relapse.⁶ Steroid-sparing immunosuppressives such as azathioprine may be used if steroids are not tolerated or ineffective.

In summary, early investigations of the unexplained neurological syndrome occurring in pork plant workers suggested the possibility of a chronic peripheral nerve disorder somehow linked to exposure to a pork brain removal process using compressed air. The common exposure to pork brain tissue certainly raised the possibility of zoonotic diseases, although likely candidates, such as *Trichinella* and prion diseases, did not seem clinically plausible. Polyneuropathy following an infectious prodrome, such as GBS, was a possibility, but the timeframe of illness better fit a chronic polyneuropathy such as CIDP. Extensive epidemiological, clinical, and laboratory investigations would eventually uncover evidence of a novel neurological syndrome with

Extensive epidemiological, clinical, and laboratory investigations would eventually uncover evidence of a novel neurological syndrome with distinct serologic and radiographic characteristics.

distinct serologic and radiographic characteristics. Details of this investigation highlight the systematic means by which investigators arrived at this unique discovery.

LATER INVESTIGATIONS

On February 8, 2008, the Centers for Disease Control and Prevention published a report in *Morbidity and Mortality Weekly Report* outlining the epidemiological, clinical, and laboratory studies of the pork

plant workers.³ The ages of the 12 patients ranged from 21 to 51 years (with a median age of 31), half were female, and all reported having been healthy before symptom onset. Patients' symptoms were predominantly sensory, involving pain, numbness, and tingling in their extremities accompanied by fatigue as well as weakness affecting mostly the lower limbs. In a few cases, weakness ranged from moderate to severe, progressing to paralysis.⁸ All patients had evidence of a polyradiculoneuropathy, with signs of nerve root irritation, and 1 patient had transverse myelitis followed by polyradiculoneuropathy. Two had early transient facial neuropathy.

Laboratory investigations excluded a battery of infections. Ten of the 12 patients demonstrated elevated CSF protein levels and spinal magnetic resonance imaging with evidence of inflammation involving peripheral nerves or nerve roots. One patient had evidence of inflammation in the anterior spinal cord. Eleven of the 12 patients had evidence of axonal or demyelinating peripheral neuropathy on electrodiagnostic tests.³ Sural nerve biopsies in 3 patients revealed mild neuropathic abnormalities and inflammation.⁸

Morbidity and Mortality Weekly Report created a working case definition using epidemiological, clinical, and diagnostic criteria (Table 1).³ The epidemiological criterion required participation in or close exposure to swine-slaughtering operations. Clinical criteria included new onset of (1) bilateral, relatively symmetric flaccid weakness or paralysis of limbs with or without cranial nerve muscles and (2) decreased or absent deep-tendon reflexes in at least the affected limbs. Diagnostic criteria included

(1) electrophysiological studies showing axonal or demyelinating peripheral neuropathy in affected limbs, (2) neuroimaging demonstrating radiculitis, myelitis, or encephalitis, and (3) CSF testing with protein levels above 45 mg/dL, with or without pleocytosis. The exclusion criterion was identification of an alternative etiology for clinical and diagnostic findings.

Case classification was stratified as confirmed, probable, or possible. A confirmed case fit the epidemiological criterion, both clinical criteria, and the first diagnostic criterion. A probable case fit the epidemiological criterion, at least 1 clinical criterion, and at least 1 diagnostic criterion. A possible case met only the epidemiological criterion and at least 1 clinical criterion. Of the 12 cases, 8 were confirmed, 2 were probable, and 2 were possible. Epidemiologically, all 12 patients reported working at or having regular contact with the head table in the pork plant, where the brain extraction procedure using compressed air took place.

All 12 patients reported working at or having regular contact with the head table in the pork plant, where the brain extraction procedure using compressed air took place.

Case control studies identified significant associations with not only working at the head table but also specifically the process of backing heads, in which the brain and remaining muscles of the head were

Table 1. *Morbidity and Mortality Weekly Report Working Case Definition for Progressive Inflammatory Neuropathy.*

Epidemiological criterion

- Participation in or close exposure to commercial or private swine-slaughtering operations.

Clinical criteria

- New onset of bilateral and relatively symmetric flaccid weakness/paralysis of the limbs, with or without involvement of cranial-nerve innervated muscles.
- New onset of decreased or absent deep-tendon reflexes, at least in affected limbs.

Diagnostic criteria

- Electrodiagnostic studies consistent with axonal or demyelinating peripheral neuropathic features in affected limbs and not attributable to an underlying chronic disease process.
- Neuroimaging consistent with radiculitis, myelitis, or encephalitis.
- Cerebrospinal fluid protein level > 45 mg/dL (with or without pleocytosis).

Exclusion criterion

- Identification of an alternative etiology for clinical or diagnostic findings.

Case classification

- Confirmed case: meets the epidemiological criterion, meets both clinical criteria, and has electrodiagnostic studies consistent with axonal or demyelinating features.
- Probable case: meets the epidemiological criterion, at least 1 clinical criterion, and at least 1 diagnostic criterion.
- Possible case: meets the epidemiological criterion and at least 1 clinical criterion.

NOTE: Table adapted from *Morbidity and Mortality Weekly Report*, February 8, 2008.³

removed with a device, placed through the pig's foramen magnum, that used compressed air to liquefy the brain to facilitate removal. Some of the workers have since stated that the process often dispersed the brain matter into the air and onto surrounding workers, raising concerns about the inhalation and absorption of the porcine neural tissue.²

Blood and throat cultures from patients did not yield any known viruses, bacteria, or parasites.³ Similarly, investigations of the pork plant environment did not yield positive tests for any evidence of microbes.² Relatives of all cases remained unaffected, and this suggested that the disease was not transmitted by human-to-human contact. Transmission also did not seem to be mediated through the ingestion of pork, as the meat processed in the plant was shipped internationally for consumption with no cases arising in those parts of the world.² Thus, the lack of evidence of transmission to persons not directly exposed to the brain extraction procedure itself greatly reduced the likelihood that the neuropathies were caused by specific infectious agents.

Focus on Progressive Inflammatory Neuropathy

The report concluded instead that the likely cause of the neuropathy was the unique exposure to pork brains in liquefied or aerosolized form, which resulted in an immune-mediated inflammatory polyneuropathy named *progressive inflammatory neuropathy* (PIN). PIN does have a variety of characteristics in common with CIDP. Both are progressive sensorimotor peripheral neuropathies with many similar diagnostic features, such as diminished deep-tendon reflexes, elevated CSF protein concentrations, and electrophysiological findings. Moreover, PIN has exhibited a chronic relapsing course. For example, the index case, as doctors refer to him, was severely affected by PIN, with significant

The report concluded instead that the likely cause of the neuropathy was the unique exposure to pork brains in liquefied or aerosolized form, which resulted in an immune-mediated inflammatory polyneuropathy named progressive inflammatory neuropathy (PIN).

lower limb weakness as well as the only documented signs to date of spinal cord inflammation.² He was treated with corticosteroids and, after a few months,

was taken off medications and returned to work; however, within weeks, symptoms returned, and he was forced to stop work and resume treatments.² Several such relapses, similar to those seen in CIDP rather than GBS, recurred.

Despite the aforementioned similarities, there are several distinguishing features of PIN. For example, unlike CIDP,⁶ PIN appears to exhibit more sensory symptoms than motor symptoms.⁸ Moreover, although CIDP has no known precipitating factors,⁷ case control studies have strongly associated the risk of PIN with a specific exposure to the pork brain extraction procedure, suggesting somehow that the inhalation or even skin exposure to the pork brain tissue may have triggered an autoimmune reaction.^{2,9}

The investigation continued in the ensuing months. Dr. Daniel Lachance, a neuroimmunologist from the Mayo Clinic, presented an update at the 60th annual meeting of the American Academy of Neurology in Chicago, IL.^{10,11} By April 2008, a total of 18 cases had been confirmed, with 13 of 15 patients tested having elevated CSF protein concentrations and showing evidence of polyradiculitis with enhancing thickened spinal roots.¹⁰ Most interestingly, investigators found that all confirmed cases of PIN displayed a novel, distinct profile of neural antibodies, including an immunoglobulin G immunostaining pattern, which, according to Dr. Lachance, had never before been seen at the Mayo Clinic's neuroimmunology laboratory.¹¹ To date, the unique pattern of antibodies has not been named, and according to Dr. Lachance, investigators are not sure exactly how many antibodies there are.⁹

Most interestingly, investigators found that all confirmed cases of PIN displayed a novel, distinct profile of neural antibodies.

FOLLOW-UP AND CURRENT STATUS

As current patients with the disease have had variable incubation periods of up to months from the time of exposure, there was potential for new cases to arise in the future. A June 5, 2008 update on the developments of investigations indicated that in addition to the 9 men and 9 women with confirmed cases of PIN in Minnesota, 5 similar cases were identified at an Indiana pork processing plant,⁹ and 1 case was identified at a Nebraska plant.⁹ Both plants also used compressed air for pork brain removal.² In an advance of the American Academy

of Neurology annual meeting in late April 2009, the Mayo Clinic researchers released an abstract of their latest findings for 24 cases ranging in age from 20 to 54. According to the update, 17 workers were treated with corticosteroids or intravenous immunoglobulin and improved, whereas 6 experienced improvement without treatment, and 1 was lost to follow-up.¹² Most of the workers have returned to work, although not in the processing area.¹²

Most of the workers have returned to work, although not in the processing area.

Mayo Clinic researchers plan to continue routine testing of workers for additional cases and to conduct investigations to determine why only some plant workers were affected. According to Dr. Lachance, the researchers do not believe that PIN had existed and gone undetected prior to the outbreak but rather had appeared during a period of accelerated processing at the plant.¹²

In conclusion, unraveling the mysterious nature of PIN continues to intrigue the medical, scientific, and epidemiological communities. For more effective treatments and better understanding of the underlying mechanism, the medical community will rely on further investigations by the teams studying this novel illness. An important step in protecting the public from this illness was the identification of the risk associated with participation in or exposure to swine slaughter and specifically the compressed air brain extraction procedure itself; this necessitated the discontinuation of the practice. As such, this illustrates the value of epidemiological principles applied by state public health officials to contain the outbreak and avoid new cases while investigators worked to uncover the pathophysiology behind this newly identified immune polyradiculoneuropathy.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Paddock C. Minnesota pork plant illness not a public risk. Available at: <http://www.medicalnewstoday.com/articles/90789.php>. Accessed June 2008.
2. Grady D. A medical mystery unfolds in Minnesota. *New York Times* February 5, 2008. Available at: <http://www.nytimes.com/2008/02/05/health/05pork.html>. Accessed June 2008.
3. Centers for Disease Control and Prevention. Investigation of progressive inflammatory neuropathy among swine slaughterhouse workers—Minnesota, 2007–2008. *MMWR Morb Mortal Wkly Rep* 2008; 57: 122–124.
4. Said G. Infectious neuropathies. *Neurol Clin* 2007; 25: 115–137.
5. Burns TM. Guillain-Barre syndrome. *Semin Neurol* 2008; 28: 152–167.
6. Saperstein DS. Chronic acquired demyelinating polyneuropathies. *Semin Neurol* 2008; 28: 168–184.
7. Said G. Chronic inflammatory demyelinating polyneuropathy. *Neuromuscul Disord* 2006; 16: 293–303.
8. Lachance DH, Dyck PJB, Pittock SJ, et al. Outbreak of immune polyradiculoneuropathy in workers exposed to porcine neural tissue. Available at: <http://www.aan.com/globals/axon/assets/4088.pdf>. Accessed June 2008.
9. Cajigal S. Antibody pattern discovered for pork plant illness. *Neurol Today* 2008; 8: 3.
10. New details presented in outbreak in pork processing plant workers. Available at: <http://www.sciencedaily.com/releases/2008/04/080416174618.htm>. Accessed June 2008.
11. Cassels C. Findings in pork workers with novel neurological illness reported for the first time. Available at: <http://www.medscape.com/viewarticle/573193>. Accessed June 2008.
12. Fitzgerald S. Condition of pork plant workers improving after developing neuropathy. *Neurol Today* 2009; 9: 18–20.

The Significant but Understudied Impact of Pathogen Transmission from Humans to Animals

Jonathan H. Epstein, DVM, MPH,¹ and Joan T. Price, BS²

¹Conservation Medicine Program, Wildlife Trust, New York, NY

²Mount Sinai School of Medicine, New York, NY

OUTLINE

ZOOANTHROPONOTIC PATHOGENS

Measles Virus and Related Paramyxoviruses

Pandemic Influenza Virus (H1N1)

Herpes Simplex Virus 1

**Methicillin-Resistant *Staphylococcus aureus*
and *Mycobacterium tuberculosis***

DISCUSSION

ABSTRACT

Zoonotic pathogens, which are transmitted from humans to nonhuman animals, are an understudied aspect of global health, despite their potential to cause significant disease burden in wild and domestic animal populations and affect global economies. Some key human-borne pathogens that have been shown to infect animals and cause morbidity and mortality include measles virus (paramyxoviruses), influenza A virus (orthomyxoviruses), herpes simplex 1 virus (herpesviruses), protozoal and helminthic parasites, and bacteria such as methicillin-resistant *Staphylococcus aureus* and *Mycobacterium tuberculosis*. However, zoonotic pathogens are most commonly reported in captive animals or domestic livestock with close human contact; there, the potential for economic loss and human reinfection is most apparent. There is also the potential for infection in wild animal populations, which may

threaten endangered species and decrease biodiversity. The emergence and reemergence of human-borne pathogens in wildlife may also have negative consequences for human health if these pathogens cycle back into humans. Many of the anthropogenic drivers of zoonotic disease emergence also facilitate zoonotic transmission. Increasing research to better understand the occurrence of and the potential for bidirectional pathogen transmission between humans and animals is essential for improving global health. *Mt Sinai J Med* 76:448–455, 2009. © 2009 Mount Sinai School of Medicine

Key Words: conservation, conservation medicine, H1N1 influenza A, herpesvirus, methicillin-resistant *Staphylococcus aureus*, paramyxovirus, primates, tuberculosis, zoonoses, zoonoses.

It is all one sickness, in the sense that our planet is suffering a systemic inflammation of *Homo sapiens*. The increasing frequency and increasing scope of zoonotic and zoonotic infections should be seen in this light—as a pattern, a set of interconnected effects, reflecting causes that are largely of human doing. —David Quammen¹

Zoonotic pathogens, which are transmitted from animals to humans, often infect persons in close contact with animals and have become an increasingly significant public health threat because of their potential to cause substantial and sometimes widespread disease in human populations.^{2,3} Furthermore, the emergence of zoonotic pathogens can inflict substantial costs on national and global economies, leading to losses in trade and tourism and increased public health care costs (Figure 1).

Less frequently reported but constituting significant global health and potential economic issues, human-borne pathogens are transmitted to nonhuman animals. Such pathogens have, in the past, been

Address Correspondence to:

Jonathan H. Epstein

Conservation Medicine Program

Wildlife Trust

New York, NY

Email: epstein@wildlifetrust.org

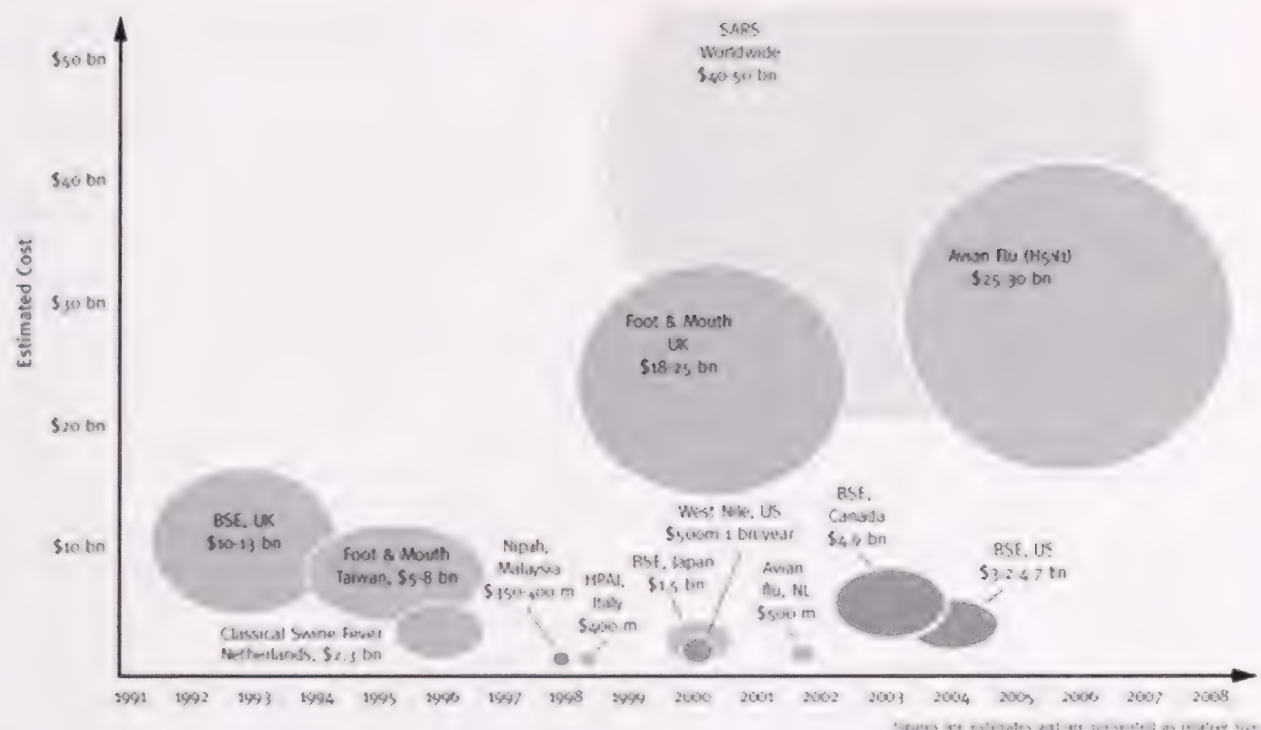


Fig 1. A chart showing the relative cost of recent outbreaks. Several of these outbreaks were due to zoonotic pathogens, including Nipah virus, SARS, West Nile virus, and avian influenza (H5N1). Figures include losses to tourism and trade and the cost of health care. **Abbreviations:** BSE, bovine spongiform encephalopathy; HPAI, highly pathogenic avian influenza; SARS, severe acute respiratory syndrome. Reprinted with permission from Bio-Era.⁽¹⁾

Human-borne pathogens that are transmitted to nonhuman animals . . . (zooanthroponoses) not only present a health burden to many animal species but may cause damage to national or global economies and threaten wild animal populations. There also may be a risk of subsequent reinfection of humans if the pathogen is able to persist in animal populations.

termed *zooanthroponoses*⁴ and not only present a health burden to many animal species but in some cases may cause damage to national or global economies and threaten wild animal populations.⁵ There also may be a risk of subsequent reinfection of humans if the pathogen is able to persist in animal populations.

ZOOANTHROPONOTIC PATHOGENS

Zooanthroponotic pathogens include viruses, bacteria, and protozoal parasites and can be transmitted whenever humans come in contact with animals, whether those animals are wildlife, livestock, or zoo or laboratory animals. These pathogens may cause diseases with varying severity depending on the species infected. For example, increased human-primate interactions through activities such as eco-tourism, research, and bushmeat hunting result in an increased risk of cross-species viral transmission.^{6,7} Nonhuman primates, including chimpanzees (*Pan troglodytes*) and gorillas (*Gorilla gorilla* spp.), have been found to be susceptible to several pathogens carried by humans, including paramyxoviruses (eg, measles virus)⁸ and intestinal bacteria,⁷ which can be detrimental to the conservation of these endangered species.⁹

Domestic animals such as cattle, goats, and sheep can perpetuate cycles of infection in humans through the contamination of drinking water with human enteric pathogens.^{10,11} *Cryptosporidium hominis* and *Giardia lamblia*, both human protozoal parasites, have been found in domestic cattle that

have access to reclaimed wastewater.¹² The clinical presentation of common zoonanthropotic infections in animals is variable, and in some cases, such as herpes simplex virus 1 (HSV-1) or measles virus in primates, the disease is often rapidly fatal. Here we review some of the major groups of zoonanthroposes, the mechanisms of transmission, and the impact on animal and human health.

Measles Virus and Related Paramyxoviruses

Humans are the natural reservoir for measles virus. Human infection clinically presents with fever, maculopapular rash, and cough with coryza and conjunctivitis.¹³ Another common sign of measles infection is Koplik's spots of the oral mucosa. The virus is transmitted via airborne droplets as well as direct contact with respiratory secretions of infected individuals.¹³ The measles virus can infect any person who has not been exposed before or who has not been successfully immunized. Both New and Old World primates, including the great apes, which are all threatened with extinction, have also been shown to be susceptible to measles and other human-related paramyxovirus virus infections in the

Both New and Old World primates, including the great apes, which are all threatened with extinction, have also been shown to be susceptible to measles and other human-related paramyxovirus virus infections in the wild.

wild.^{14,15} Great apes can become exposed to human viruses through close contact with tourists, researchers, or poachers.¹⁵ Several outbreaks of respiratory illness were observed in chimpanzees at a research site in Côte d'Ivoire between 1999 and 2006.¹⁴ Morbidity rates were 92%, and in 3 outbreaks that involved mortality, the rates were 13% on average ($n = 110$) and were as high as 19% ($n = 32$). Human metapneumovirus and human respiratory syncytial virus are both related to measles virus (family *Paramyxoviridae*) and are common respiratory illnesses in Africa.^{16,17} Human metapneumovirus and human respiratory syncytial virus were found in some of the dead chimpanzees, and phylogenetic analyses indicated that these were of human origin.¹⁴

Rhesus macaques (*Macaca mulatta*), which are often used as human models in biomedical research because of their physiological response to many human pathogens, have been shown to have a pathology similar to that of humans when infected with measles. In cases of natural infection in a wild setting and in cases of infection in a laboratory, they experience a progression of clinical signs very similar to those experienced by humans, including maculopapular rash, fever, conjunctivitis, and even Koplik's spots on mucosal membranes.^{13,18} As with humans, macaques who recover from an initial measles infection will demonstrate immunity to measles virus.¹⁸

Measles infections in nonhuman primates may also have varied or more severe courses. In one experiment, it was found that infected marmosets (*Callithrix* spp.) lacked symptoms commonly found in humans and some nonhuman primates, such as rash, coryza, and conjunctivitis, and they exhibited a longer subclinical infectious period and a much higher mortality rate.¹⁹ The lack of rash and the increased mortality among these animals have been attributed to a decrease in the immunological response to the measles virus.¹⁹ A longer subclinical period may make infected marmosets more dangerous to susceptible humans.

In one outbreak that occurred among Japanese macaques (*Macaca fuscata*) in a Korean zoo, the source could not be identified, but it seemed that the monkeys were infected by aerosol from infected visitors.²⁰ Outbreaks of measles in primates have been shown to occur concurrently with human outbreaks of the same virus type.^{20,21} In this case, the Japanese macaques suffered many secondary infections not normally seen in infected immunocompetent adult humans, and this indicated marked immunosuppression among infected primates.

Vaccination of wildlife or other animals at risk of exposure, where feasible, may be a beneficial strategy for preventing infection. Measles vaccine is highly effective in humans. One dose of live attenuated measles vaccine provides active immunity in 94% to 98% of susceptible persons, whereas a second dose will induce immunity levels of up to 99%.¹³ Immunization with measles vaccine has been shown to be effective in some primate species,^{19,22} and captive gorillas showed persistent immunity for up to 11 years post-inoculation. The potential for high mortality rates in endangered species may warrant immunization against measles in both captive and free-ranging populations that have high levels of contact with humans. Furthermore, improved regulation that limits contact between tourists and researchers and wild or laboratory primates may also

reduce the risk of zoonanthroponotic transmission of pathogens.⁷

Pandemic Influenza Virus (H1N1)

Wild birds are considered the natural reservoir for influenza A viruses (family *Orthomyxoviridae*) and have been shown to carry all hemagglutinin (H) and neuraminidase (N) subtypes.²³ Humans and other mammals, including pigs, are susceptible to infection with certain subtypes of influenza A virus, including pandemic strains of H1N1 (the etiological agent of both the 1918 Spanish flu and 2009 pandemic). Humans are susceptible to only 3 types of flu, H1N1, H1N2, and H2N3, all of which originated in birds but

Humans and other mammals, including pigs, are susceptible to infection with certain subtypes of influenza A virus, including pandemic strains of H1N1 (the etiological agent of both the 1918 Spanish flu and 2009 pandemic). . . . Pigs are unique in that they are susceptible to infection with both avian and human subtypes of influenza A; this allows them to act as mixing vessels for new strains via genetic reassortment.

have become endemic to human populations and circulate globally. H5N1 remains an avian subtype to which people are highly susceptible, but it cannot easily transmit to other people. Pigs are unique in that they are susceptible to infection with both avian and human subtypes of influenza A; this allows them to act as mixing vessels for new strains via genetic reassortment.²⁴ Pigs are likely to have been infected with the pandemic H1N1 strain of 1918 by circulating human strains after the first wave of infection swept through the United States in 1918.²⁵ The pandemic in 1918 was the first instance of flu being described in pigs, and it spread widely throughout the United States, leading to recurrent outbreaks in pig populations and the name *swine flu*. H1N1 causes varying levels of morbidity in pigs, although it is rarely fatal. Strains of H1N1 have circulated in domestic swine herds in North America and Europe for decades since 1918, although in the 1990s, H2N3 became the predominant strain in pigs in North America.²⁶⁻²⁸

In 2009, a novel strain of H1N1 emerged in Veracruz, Mexico and quickly spread to the United States and more than 30 other countries via global travel networks, rapidly becoming a global pandemic.²⁷ Genomic sequencing of the 2009 pandemic H1N1 strain revealed that it was a novel strain that included genetic material with avian, swine, and human origins but that was not known to be circulating in swine herds.²⁹ The usual economic impact of H1N1 infection in swine herds is due to the high levels of morbidity, which include substantial weight loss and restriction on movement. However, H1N1 had a significant impact on the Mexican economy, which had an estimated \$2.2 billion loss due to the closure of businesses and a drop in travel and tourism,³⁰ as well as a negative effect on the greater global economy because of the incorrect presumption that pigs were responsible for the human cases.^{31,32} Additionally, the false perception that consuming pork could cause infection further affected the global swine trade.³³

The only confirmed outbreak to date of the novel H1N1 strain in pigs occurred in a swine herd in Canada in March 2009. A laborer from the farm, who had recently traveled to Mexico, returned sick and may have introduced the H1N1 virus to the herd.³⁴ As the number of human cases increases globally, so too does the risk of additional instances of zoonanthroponotic flu transmission, as was the case in 1918.

Herpes Simplex Virus 1

Herpes viruses as a group are associated with high mortality rates when exchange occurs between humans and nonhuman primates.³⁵⁻³⁸ HSV-1 infection in humans is usually unapparent but may cause mild, recurrent oral and facial lesions.¹³ More severe symptoms such as fever and malaise, keratoconjunctivitis, and pharyngotonsillitis are present in some 10% of primary infections, particularly in neonates and immunocompromised adults. Central nervous system involvement may also occur, although far less frequently. HSV-1 can be transmitted in the saliva of both symptomatic and asymptomatic humans. Humans are the natural reservoir for the HSV-1 virus, and 50% to 90% of adults worldwide have antibodies against it.

Although herpes simplex viruses cause relatively mild disease in immune-competent humans, they can be fatal when spread to many species of nonhuman primates, including gorillas (*Gorilla gorilla*), patas monkeys (*Erythrocebus patas*), colobus monkeys (*Colobus* spp.), marmosets (*Callithrix jacchus*),

*Although herpes simplex viruses cause relatively mild disease in immune-competent humans, they can be fatal when spread to many species of nonhuman primates, including gorillas (*Gorilla gorilla*), patas monkeys (*Erythrocebus patas*), colobus monkeys (*Colobus spp.*), marmosets (*Callithrix jacchus*), and white-handed gibbons (*Hylobates lar*).*

and white-handed gibbons (*Hylobates lar*).^{35–38} Most animals infected with HSV-1 present with oral vesicles, as in human hosts, but the disease usually progresses to more severe sequelae, including conjunctivitis, meningoencephalitis, and death.³⁹ The most common outbreaks of HSV-1 in nonhumans have been described in zoo animals. In one such outbreak, 3 white-faced saki monkeys (*Pithecia pithecia*) in a zoo exhibit developed symptoms such as nasal discharge, anorexia, dehydration, fever, and seizures as well as oral ulcers.⁴⁰ All 3 died within 48 to 96 hours of the onset of the initial signs of disease. Necropsies revealed oral and esophageal lesions, liver lesions, and meningoencephalitis. The origin of the infection was thought to be visitors or a zookeeper. Researchers suspected that 1 monkey was initially infected, and then the virus quickly spread to the other 2 monkeys. The rapid transmission of HSV-1 to these monkeys and their death within 3 days suggest a very high susceptibility and illustrate the need for greater precautions to prevent zoonanthroponotic infection in zoo settings.

Although HSV-1 causes acute fatal disease in a number of nonhuman primates, it does spare some species such as the macaque (*Macaca spp.*).⁴¹ Researchers believe that the reason macaques are less susceptible to HSV-1 infection than other primates is that they are the natural host of another herpes simplex virus called herpes B virus, or cercopithecine-1.⁴¹ Herpes B virus, like HSV-1 in nonhuman primates, causes acute fatal disease when people are infected. Although there have been substantial efforts toward research and prevention of herpes B in primate handlers in zoo and laboratory settings, recent studies of HSV-1 infection among captive animals remain scarce.^{35–38}

Methicillin-Resistant *Staphylococcus aureus* and *Mycobacterium tuberculosis*

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a bacterial pathogen that has historically been the leading cause of hospital-based infections globally.⁴² Recently, new strains of community-acquired MRSA have been identified and are an emerging public health threat and across the United States.⁴³ MRSA typically causes severe skin and soft tissue damage, and recently, community-acquired MRSA has also been identified as a cause of necrotizing pneumonia.⁴⁴ MRSA has been

MRSA has been reported in several livestock and companion animal species, and transmission between animals and between humans and animals has been described.

reported in several livestock and companion animal species, and transmission between animals and between humans and animals has been described.⁴⁵ Companion animals are vulnerable to hospital-acquired MRSA infections but may also acquire MRSA environmentally if living in a household with an infected person.⁴⁵ Long-term health care facilities have a high incidence of MRSA infection in their patients, who carry it from acute care facilities.⁴⁶ Interestingly, a recent study by Lefebvre *et al.*⁴⁷ showed that dogs that visited human health care facilities were more likely to become infected with MRSA than those that did not. These dogs were likely to have acquired the infection by licking patients or accepting treats from them. These dogs also moved from clinic to clinic, creating a potential opportunity to act as a vector for MRSA.

Humans are the natural reservoir for the bacterium *Mycobacterium tuberculosis*.¹³ An infection from *M. tuberculosis* can have long periods of latency before proceeding to active tuberculosis (TB), which most commonly causes infiltrates and cavitations of the lungs, although extrapulmonary infections can be found in any tissue, including the lymph nodes, the central nervous system, kidneys, bones, and joints. Humans infected with TB may present initially with fever, fatigue, night sweats, and weight loss during the early period of infection and then with cough, chest pain, and hemoptysis as the disease progresses. TB can often be fatal, especially in immunocompromised hosts or when left untreated.¹³

Mycobacterium tuberculosis can also cause disease in a wide variety of animal species, including

primates, elephants, Rocky Mountain goats, black rhinoceroses, and marine mammals.^{48–50} As in humans, most animals develop pulmonary disease when infected with TB. Parrots present with skin and mucous membrane lesions.^{51,52}

Anthropozoonotic transmission of *Mycobacterium tuberculosis* has been demonstrated in monkeys,^{50,53} cattle,⁵⁴ parrots,^{51,52} and elephants,⁴⁸ mostly in captive settings. Zoonotic transmission of

*Anthropozoonotic transmission of
Mycobacterium tuberculosis has
been demonstrated in monkeys,
cattle, parrots, and elephants,
mostly in captive settings.*

TB has also been documented.⁴⁹ TB in animals can often be subclinical for longer periods of time in comparison with humans. Tuberculin tests, similar to those done in humans, can effectively detect subclinical disease in animals.⁵⁵ It has been suggested that humans may be at risk for TB infection when in contact with infected animals such as pet parrots; however, it is not known whether birds can reinfect humans.⁵²

DISCUSSION

Zooanthroponoses negatively affect the health of livestock, companion animals, free-ranging wildlife, and zoo animals and can threaten populations of endangered species.¹⁴ Zoonotic pathogens are responsible for 75% of emerging infectious diseases in humans, and outbreaks of zoonotic pathogens can have significant economic consequences globally (Figure 1).⁵⁶ Thus, it is not surprising that substantially more resources have been devoted to zoonotic disease research rather than zooanthroponotic disease research. One result of this, however, has been a lack of adequate surveillance in animals for human-borne pathogens, which may also have significant health and economic consequences.

Routine disease surveillance and diagnostic testing in animals are primarily reserved for large-scale livestock productions, for which there are economic incentives to maintain herd health, or for companion animals in countries in which owners are willing and able to pay for it. Infectious disease surveillance and control in wild animal populations have been much less common and typically occur in the context of a known zoonotic agent (eg, H5N1 and wild birds) or perhaps in endangered species.

Unfortunately, the economic incentives for detecting and controlling infectious disease in wildlife are less obvious than in domestic animals, and in areas in which zooanthroponotic transmission is of high conservation importance (eg, in African countries with many of the great ape species), local wildlife authorities often do not have the necessary resources to identify etiological agents in an animal outbreak. Even though many zooanthroponotic infections occur in captive animals because of contact with researchers, trainers, and keepers, these pathogens may potentially emerge in free-living populations of the same or related species and have far-reaching detrimental effects on biodiversity and human health.² Furthermore, surveillance and control measures designed to prevent zooanthroponotic infections may also serve to protect human health by preventing spillback from animals to humans. In the case of flu, the infection of pigs with human influenza viruses has the additional risk of creating novel, potentially more virulent flu strains via genetic reassortment.⁵⁷

The transmission of pathogens between humans and nonhuman species (in either direction) is driven by anthropogenic factors that increase contact between humans and animals.^{2,58} These factors include agricultural expansion and intensification,

*The transmission of pathogens
between humans and nonhuman
species (in either direction) is
driven by anthropogenic factors
that increase contact between
humans and animals. These
factors include agricultural
expansion and intensification,
global travel (tourism, business,
and emigration), animal trade,
and urbanization, all of which
are likely to increase as the human
population continues to grow.*

global travel (tourism, business, and emigration), animal trade, and urbanization, all of which are likely to increase as the human population continues to grow.^{3,58} Globally, regions with high biodiversity, human population density, and high levels of human-animal interaction are predicted to be most vulnerable to the emergence of zoonotic pathogens.³ The same criteria are also likely to apply to zooanthroponotic pathogens.

The health and economic impacts of zoonotic diseases are now more widely appreciated since the severe acute respiratory syndrome coronavirus and highly pathogenic avian influenza (H5N1) outbreaks (Figure 1). Although it is expected that there will continue to be a research focus on zoonotic pathogens, it would seem prudent to allocate more resources to screening wildlife and domestic animals for human pathogens and to studying the epidemiology of these pathogens as they cycle between humans and animals. Comprehensive scientific disciplines, such as conservation medicine, that recognize the links between human, animal, and ecosystem health, provide an effective approach to understanding the complex multidirectional exchange of pathogens among humans, wildlife, and domestic animals.^{2,59} As the rate of emerging infectious diseases (in humans and animals) is likely to increase with an increase in anthropogenic pressures, it is important that we develop a better understanding of the potential for and occurrence of cross-species pathogen transmission in order to successfully protect human, animal, and ecosystem health.

ACKNOWLEDGMENT

This work was supported by the National Institute of Allergy and Infectious Diseases (grant K08AI067549 to Jonathan H. Epstein).

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Quammen D. Little is big, many is one: zoonoses in the twenty-first century. In: Eva Fearn and Ward Woods Editors eds. *State of the Wild: A Global Portrait of Wildlife, Wildlands, and Oceans 2008–2009*. Washington, DC: Island Press; 2008.
- Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* 2000; 287: 443–449.
- Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature* 2008; 451: 990–993.
- Hubalek Z. Emerging human infectious diseases: anthroponoses, zoonoses, and sapronoses. *Emerg Infect Dis* 2003; 9: 403–404.
- Wolfe ND, Daszak P, Kilpatrick AM, Burke DS. Bushmeat hunting, deforestation and prediction of zoonotic emergence. *Emerg Infect Dis* 2005; 11: 1822–1827.
- Goldberg TL, Gillespie TR, Rwego IB, et al. Forest fragmentation and bacterial transmission among nonhuman primates, humans, and livestock, Uganda. *Emerg Infect Dis* 2008; 14: 1375–1382.
- Goldberg TL, Gillespie TR, Rwego IB, et al. Patterns of gastrointestinal bacterial exchange between chimpanzees and humans involved in research and tourism in western Uganda. *Biol Conserv* 2007; 135: 511–517.
- Kaur T, Singh J, Tong SX, et al. Descriptive epidemiology of fatal respiratory outbreaks and detection of a human-related metapneumovirus in wild chimpanzees (*Pan troglodytes*) at Mahale Mountains National Park, western Tanzania. *Am J Primatol* 2008; 70: 755–765.
- Wallis J. Primate conservation and human to nonhuman primate disease transmission. *Am J Phys Anthropol* 2003; suppl. 36: 219–219.
- Tzipori S. Cryptosporidiosis in animals and humans. *Microbiol Rev* 1983; 47: 84–96.
- Lobo ML, Xiao L, Antunes F, Matos O. Occurrence of *Cryptosporidium* and *Giardia* genotypes and subtypes in raw and treated water in Portugal. *Lett Appl Microbiol* 2009; 48: 732–737.
- Di Giovanni GD, Betancourt WQ, Hernandez J, et al. Investigation of potential zooanthroponotic transmission of cryptosporidiosis and giardiasis through agricultural use of reclaimed wastewater. *Int J Environ Health Res* 2006; 16: 405–418.
- Heymann D, ed. *Control of Communicable Diseases Manual*. 18th ed. Washington, DC: American Public Health Association; 2004.
- Kondgen S, Kuhl H, N'Goran PK, et al. Pandemic human viruses cause decline of endangered great apes. *Curr Biol* 2008; 18: 260–264.
- Woodford MH, Butynski TM, Karesh WB. Habituating the great apes: the disease risks. *Oryx* 2002; 36: 153–160.
- Zar HJ. Pneumonia in HIV-infected and HIV-uninfected children in developing countries: epidemiology, clinical features, and management. *Curr Opin Pulm Med* 2004; 10: 176–182.
- Weber MW, Mulholland EK, Greenwood BM. Respiratory syncytial virus infection in tropical and developing countries. *Trop Med Int Health* 1998; 3: 268–280.
- Blake FG, Trask JD Jr. Studies on measles: II. Symptomatology and pathology in monkeys experimentally infected. *J Exp Med* 1921; 33: 413–422.
- Albrecht P, Lorenz D, Klutch MJ, et al. Fatal measles infection in marmosets pathogenesis and prophylaxis. *Infect Immun* 1980; 27: 969–978.
- Choi YK, Simon MA, Kim DY, et al. Fatal measles virus infection in Japanese macaques (*Macaca fuscata*). *Vet Pathol* 1999; 36: 594–600.
- Willy ME, Woodward RA, Thornton VB, et al. Management of a measles outbreak among Old World nonhuman primates. *Lab Anim Sci* 1999; 49: 42–48.
- Blasier MW, Travis DA, Barbiers R. Retrospective evaluation of measles antibody titers in vaccinated captive gorillas (*Gorilla gorilla gorilla*). *J Zoo Wildl Med* 2005; 36: 198–203.
- Webster RG, Bean WJ, Gorman OT, et al. Evolution and ecology of influenza-A viruses. *Microbiol Rev* 1992; 56: 152–179.

24. Ito T, Couceiro J, Kelm S, et al. Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J Virol* 1998; 72: 7367–7373.
25. Shope RE. The incidence of neutralizing antibodies for swine influenza virus in the sera of human beings of different ages. *J Exp Med* 1936; 63: 669–684.
26. Dunham EJ, Dugan VG, Kaser EK, et al. Different evolutionary trajectories of European avian-like and classical swine H1N1 influenza A viruses. *J Virol* 2009; 83: 5485–5494.
27. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 2009; 459: 931–939.
28. Olsen CW. The emergence of novel swine influenza viruses in North America. *Virus Res* 2002; 85: 199–210.
29. Dawood FS, Jain S, Finelli L, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med* 2009; 360: 2605–2615.
30. Mexico estimates flu cost economy \$2.2 billion. *Wall Street Journal*. Available at: <http://online.wsj.com/article/sb124152252457387033.html>. Accessed June 2009.
31. Bryson JH. Economic effects of swine flu: Mexico and beyond. Available at: <http://www.wachovia.com/economics>. Accessed 30 June 2009.
32. Global development finance report. Available at: http://siteresources.worldbank.org/extavianflu/resources/gdf2009_h1n1_impact.pdf. Accessed June 2009.
33. Cima G. H1N1 strain in humans causes concerns over animal impacts. *J Am Vet Med Assoc* 2009; 234: 1358–1360.
34. World Organisation for Animal Health. A/H1N1 influenza, Canada (immediate notification). Available at: http://www.oie.int/wahis/public.php?page=weekly_report_index&admin=0. Accessed June 2009.
35. Heldstab A, Ruedi D, Sonnabend W, Deinhardt F. Spontaneous generalized *Herpesvirus hominis* infection of a lowland gorilla (*Gorilla gorilla gorilla*). *J Med Primatol* 1981; 10: 129–135.
36. Loomis MR, O'Neill T, Bush M, Montali RJ. Fatal herpesvirus infection in patas monkeys and a black and white colobus monkey. *J Am Vet Med Assoc* 1981; 179: 1236–1239.
37. Matz-Rensing K, Jentsch KD, Rensing S, et al. Fatal herpes simplex infection in a group of common marmosets (*Callithrix jacchus*). *Vet Pathol* 2003; 40: 405–411.
38. Ramsay E, Stair EL, Castro AE, Marks MI. Fatal *Herpesvirus hominis* encephalitis in a white-handed gibbon. *J Am Vet Med Assoc* 1982; 181: 1429–1430.
39. McClure HM, Keeling ME. Viral diseases noted in the Yerkes Primate Center colony. *Lab Anim Sci* 1971; 21: 1002–1010.
40. Schrenzel MD, Osborn KG, Shima A, et al. Naturally occurring fatal herpes simplex virus 1 infection in a family of white-faced saki monkeys (*Pithecia pithecia*). *J Med Primatol* 2003; 32: 7–14.
41. Weigler BJ. Biology of B virus in macaque and human hosts: a review. *Clin Infect Dis* 1992; 14: 555–567.
42. Oliveira DC, Tomasz A, de Lencastre H. Secrets of success of a human pathogen: molecular evolution of pandemic clones of methicillin-resistant *Staphylococcus aureus*. *Lancet Infect Dis* 2002; 2: 180–189.
43. Zetola N, Francis JS, Nuermberger EL, Bishai WR. Community-acquired methicillin-resistant *Staphylococcus aureus*: an emerging threat. *Lancet Infect Dis* 2005; 5: 275–286.
44. Hidron AI, Low CE, Honig EG, Blumberg HM. Emergence of community-acquired methicillin-resistant *Staphylococcus aureus* strain USA300 as a cause of necrotising community-onset pneumonia. *Lancet Infect Dis* 2009; 9: 384–392.
45. Morgan M. Methicillin-resistant *Staphylococcus aureus* and animals: zoonosis or humanosis? *J Antimicrob Chemother* 2008; 62: 1181–1187.
46. Crossley K. Long-term care facilities as sources of antibiotic-resistant nosocomial pathogens. *Curr Opin Infect Dis* 2001; 14: 455–459.
47. Lefebvre SL, Reid-Smith RJ, Waltner-Toews D, Weese JS. Incidence of acquisition of methicillin-resistant *Staphylococcus aureus*, *Clostridium difficile*, and other health-care-associated pathogens by dogs that participate in animal-assisted interventions. *J Am Vet Med Assoc* 2009; 234: 1404–1417.
48. Michalak K, Austin C, Diesel S, et al. *Mycobacterium tuberculosis* infection as a zoonotic disease: transmission between humans and elephants. *Emerg Infect Dis* 1998; 4: 283–287.
49. Montali RJ, Mikota SK, Cheng LI. *Mycobacterium tuberculosis* in zoo and wildlife species. *Rev Sci Tech* 2001; 20: 291–303.
50. Moreland AF. Tuberculosis in New World primates. *Lab Anim Care* 1970; 20: 262–264.
51. Ackerman LJ, Benbrook SC, Walton BC. *Mycobacterium tuberculosis* infection in a parrot (*Amazona farinosa*). *Am Rev Respir Dis* 1974; 109: 388–390.
52. Washko RM, Hofer H, Kiehn TE, et al. *Mycobacterium tuberculosis* infection in a green-winged macaw (*Ara chloroptera*): report with public health implications. *J Clin Microbiol* 1998; 36: 1101–1102.
53. Michel AL, Huchzermeyer HF. The zoonotic importance of *Mycobacterium tuberculosis*: transmission from human to monkey. *J S Afr Vet Assoc* 1998; 69: 64–65.
54. Ocepek M, Pate M, Zolnir-Dovc M, Poljak M. Transmission of *Mycobacterium tuberculosis* from human to cattle. *J Clin Microbiol* 2005; 43: 3555–3557.
55. Oh P, Granich R, Scott J, et al. Human exposure following *Mycobacterium tuberculosis* infection of multiple animal species in a Metropolitan Zoo. *Emerg Infect Dis* 2002; 8: 1290–1293.
56. Taylor LH, Latham SM, Woolhouse ME. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci* 2001; 356: 983–989.
57. Alexander DJ, Brown IH. Recent zoonoses caused by influenza A viruses. *Rev Sci Tech* 2000; 19: 197–225.
58. Daszak P, Cunningham AA, Hyatt AD. Anthropogenic environmental change and the emergence of infectious diseases in wildlife. *Acta Trop* 2001; 78: 103–116.
59. Daszak P, Tabor GM, Kilpatrick AM, et al. Conservation medicine and a new agenda for emerging diseases. In: Bokma BH, Blouin E, Bechara G.H. eds. *Impact of Ecological Changes on Tropical Animal Health and Disease Control*. New York, NY: New York Academy of Sciences; 2004; 1–11.
60. Bio-Era. Available at: <http://www.bio-era.net/activities/thinking.html>. Accessed 30 June 2009.

The Plague of Athens: Epidemiology and Paleopathology

Robert J. Littman, MLitt, PhD

Department of Languages and Literatures of Europe and the Americas,
University of Hawaii, Honolulu, HI

OUTLINE

THE PLAGUE OF ATHENS: THUCYDIDES' ACCOUNT
THE PLAGUE OF ATHENS: ANALYSES IN THE 20TH CENTURY
EPIDEMIOLOGICAL INTERPRETATION

Person

Place

Time

Other Epidemiological Information

Thucydides' Text

Common Source Acquisition

Person-to-Person Transmission

Mathematical Modeling

Discussion

Respiratory Transmission

Reservoir Transmission

PALEOPATHOLOGY

ABSTRACT

In 430 BC, a plague struck the city of Athens, which was then under siege by Sparta during the Peloponnesian War (431–404 BC). In the next 3 years, most of the population was infected, and perhaps as many as 75,000 to 100,000 people, 25% of the city's population, died. The Athenian general and historian Thucydides left an eye-witness account of this plague and a detailed description to allow future generations to identify the disease should it break out again. Because of the importance of Thucydides and Athens in Western history and culture, the Plague of Athens has taken a prominent position in the history of the West for the past 2500 years.

Despite Thucydides' careful description, in the past 100 years, scholars and physicians have disagreed about the identification of the disease. Based on clinical symptoms, 2 diagnoses have dominated the modern literature on the Athenian plague: smallpox and typhus. New methodologies, including forensic anthropology, demography, epidemiology, and paleopathology, including DNA analysis, have shed new light on the problem. Mathematical modeling has allowed the examination of the infection and attack rates and the determination of how long it takes a disease to spread in a city and how long it remains endemic. The highly contagious epidemic exhibited a pustular rash, high fever, and diarrhea. Originating in Ethiopia, it spread throughout the Mediterranean. It spared no segment of the population, including the statesman Pericles. The epidemic broke in early May 430 BC, with another wave in the summer of 428 BC and in the winter of 427–426 BC, and lasted 4.5 to 5 years. Thucydides portrays a virgin soil epidemic with a high attack rate and an unvarying course in persons of different ages, sexes, and nationalities.

The epidemiological analysis excludes common source diseases and most respiratory diseases. The plague can be limited to either a reservoir diseases (zoonotic or vector-borne) or one of the respiratory diseases associated with an unusual means of persistence, either environmental/fomite persistence or adaptation to indolent transmission among dispersed rural populations. The first category includes typhus, arboviral diseases, and plague, and the second category includes smallpox. Both measles and explosive streptococcal disease appear to be much less likely candidates.

In 2001, a mass grave was discovered that belonged to the plague years. Ancient microbial typhoid (*Salmonella enterica* serovar Typhi) DNA was extracted from 3 skeletons. Because typhoid was endemic in the Greek world, it is not the likely cause of this sudden epidemic. ***Mt Sinai J Med* 76:456–467, 2009. © 2009 Mount Sinai School of Medicine**

Key Words: epidemic, Plague of Athens, smallpox, typhus.

Address Correspondence to:

Robert J. Littman

Department of Languages and
Literatures of Europe and the
Americas University of Hawaii
Honolulu, HI

Email: littman@hawaii.edu

Health and disease have played an important part in human religion and history. Although our conquest of disease has extended the modern lifespan to 78 years in the Western world versus 25 to 35 years in the ancient world, we are still frightened by and concerned with plagues. In the modern world, emerging strains of viruses and bacteria, such as influenza and acquired immune deficiency syndrome, cause much anxiety and millions of deaths on a yearly basis. These continuing tragedies are reminiscent of past centuries in which a husband and wife endured the deaths of half their children from infectious disease.¹ In addition to the normal patterns of disease in childhood, epidemics would often strike out of nowhere and carry off large percentages of the population. For example, a plague that occurred in the reign of Marcus Aurelius (AD 161–180), which was described extensively by Galen, killed a tenth of the population of the Roman Empire.^{2,3} The bubonic plague struck in Byzantium in the 6th century and killed a substantial part of the city's population. This plague, named the Plague of Justinian after the sitting emperor, was extensively described in a contemporary account, modeled on the work of Thucydides, by the historian Procopius in *History of the Wars* (II 22–33). The worst plague in human history was the Black Death, which wiped out nearly half the population of Western Europe in the 14th century.^{4,5} A contemporary literary account of the effects on society survives in Boccaccio's *Decameron*. Another famous literary account of a plague occurs in Daniel Defoe's *Journal of the Plague Year*, a semifictional account of the bubonic plague in London in the 17th century (1665).

Over the past 2000 years, we have learned much about infectious diseases, the existence of microorganisms, their causes, their symptoms, and how they are spread. Since its beginning, Western medicine has concentrated on the clinical aspects of disease. In the 19th century, epidemiology began to emerge as a scientific discipline, although its roots go all the way back to Hippocrates in the 5th century BC.^{6–8} This new discipline has proved invaluable in learning the causes of diseases, thereby aiding modern medicine in formulating cures.

Despite some overlap, clinical medicine and epidemiology are distinct. Clinical medicine is the care and study of sick individuals, the examination of signs and symptoms present in sick persons, and attempts to treat the patient. Epidemiology has often been called the study of the distribution and determinants of disease in human populations; it seeks to determine who is infected, when, where, and why. It questions whether a disease strikes the young, the old, the sexually active, or those who

smoke. Epidemiology, unlike clinical medicine, studies groups of people, not individuals. It also studies the well because those who do not contract a disease can often provide as many clues as those who do. A more modern definition of epidemiology is the study of the distribution and determinants of health events, including diseases, in human populations.

Ultimately, of course, the true test of public health, of which epidemiology is a component, is whether it can help us prevent disease. Arcane knowledge and intellectual formulations are of little use when the litmus test of lives saved and health improved stands visibly before us. Epidemiology has been less than successful in some of its applications. There is and perhaps always has been a tendency for epidemiology to become intellectualized, in part because of its complex and philosophical underpinnings.

However, consideration of the epidemiological approach with respect to the Athenian epidemic will have, I hope, some modest by-products. Among these are lessons in the approach to public health problem solving that can be applied to many new diseases that exist today and will be encountered in the future. What is needed is not only more technical knowledge about agents, hosts, and environments but also perspectives on and formulae for understanding the complex ways in which they interact, structures for assembling and synthesizing disparate facts, and approaches for deriving conclusions. Although this will not, in and of itself, create a healthier world, and the derived conclusions may not motivate human beings to change the behavioral risk components

Although modern epidemiologists and physicians are most concerned with current disease, nonetheless it can also be valuable to examine the diseases of history, both to understand long-term sociological and demographic changes and to better understand how diseases work over centuries.

over which they theoretically have some control, it may at least assist us in creating a framework and a database structure from which to proceed in making public health decisions.

Although modern epidemiologists and physicians are most concerned with current disease, nonetheless it can also be valuable to examine the diseases of history, both to understand long-term

sociological and demographic changes and to better understand how diseases work over centuries.⁹

THE PLAGUE OF ATHENS: THUCYDIDES' ACCOUNT

One disease that has particularly attracted the attention of both laymen and scholars alike is the Plague of Athens. Although this was not the most important plague in history in terms of mortality and political and socioeconomic consequences, the place of the 5th century BC Athenian historian, Thucydides, as one of the fathers of history has made his description one of the best known passages in Western literature and the history of medicine.

Thucydides presents a detailed description of the symptoms of the Plague of Athens that is marked by careful observation and is woven into a terse narrative about the devastation of war and disease. In his *History of the Peloponnesian War* (2.47–55), Thucydides¹⁰ describes how the disease was accompanied by a pustular rash, high fever, and diarrhea and how it usually resulted in the death of the victim.

In his History of the Peloponnesian War (2.47–55), Thucydides describes how the disease was accompanied by a pustular rash, high fever, and diarrhea and how it usually resulted in the death of the victim.

He relates that it originated in Ethiopia and spread into Egypt and Libya and through the Near East before arriving at the Piraeus, the port of Athens. From the Piraeus, the epidemic moved rapidly into the city itself. Thucydides indicates that it was highly contagious and infected anyone who cared for the sick. He himself suffered from the plague but was one of the lucky ones who survived. Others were not so fortunate. Entire families were wiped out, and the dead became so numerous that corpses could not be buried and were left where they had died.

Thucydides (2.48.3) states that he will set down the nature of the disease and explain the symptoms by which it may be recognized if it should ever break out again. He was influenced by Hippocrates,^{11,12} who lived at the same time, both in his description of the plague and also in his application of Hippocratic doctrine to his writing of history. In a sense, his account of the disease parallels the overall themes

of his work: history as an example and a force that repeats and the decay of Athens by both war and disease.

The political ramifications of the Plague of Athens were enormous because it struck Athens shortly after the Peloponnesian War (431–404 BC) between Athens and Sparta had begun (430 BC). Although Athens was devastated by the outbreak, Sparta was relatively unscathed because the epidemic did not penetrate the Spartan homeland in the Peloponnese. The Plague of Athens carried off perhaps 75,000 to 100,000 people, about 25% of the population of Athens during the first few years of the Peloponnesian War. The Athenian leader Pericles perished of the disease in 429 BC. The plague was a major reason for the defeat of Athens in the Peloponnesian War. It was economically and socially devastating to Athens, both at the time and in the subsequent centuries. The combination of disease and war depopulated Athens and changed Greek history, which might have been very different had Athens won the war.

THE PLAGUE OF ATHENS: ANALYSES IN THE 20TH CENTURY

Numerous essays on the Plague of Athens appeared in the 20th century, attempting to identify the Athenian pestilence with 1 or more diseases. Because Thucydides provided the symptoms, the resultant studies were symptomatologies, which were correlated with a disease that presumably matched the ancient description. Two diagnoses became prominent in the modern literature on the Athenian plague. The first diagnosis was smallpox, which seemed to incorporate almost all of the Thucydidean details.¹³

The second most common retrospective diagnosis has been endemic typhus, as represented by Hans Zinsser,¹⁴ who was one of the pioneers in the study of typhus and developed the first typhus vaccine in 1934. Two other frequently cited studies followed Zinsser's diagnosis, those of MacArthur¹⁵ and Scarborough.¹⁶ Scarborough especially called attention to Thucydides' description as accurate in its own context but incapable of accounting for any evolutionary shifts in the causative organisms. He argued that any modern diagnosis is flawed because of possible mutations over time. Although Scarborough might be correct if the disease described by Thucydides was caused by a microorganism that was prone to mutation (as typhus is), other microorganisms, such as smallpox, are very stable over time.

In 1994, Thomas E. Morgan¹⁷ analyzed Thucydides' medical knowledge, or the lack of it. A

physician-classicist, Morgan argued that Thucydides desired to contrast the tragedy of war and the pathos of the disease with the lofty ideals presented by Pericles in the funeral oration. Thus, in his opinion, the descriptions of the plague are quite imprecise not only because the Hippocratic or other terminologies varied from those of modern epidemiology but also, more importantly, because Thucydides employed dramatic license in his account of the plague. The weakness in Morgan's argument is that his thesis is somewhat inconsistent with Thucydides' statement that he describes the symptoms so that the disease can be identified if it breaks out again.

Other possible candidates for the identification of the Plague of Athens have ranged from measles, typhoid fever, bubonic plague, anthrax, and ergotism to toxic shock syndrome with *Staphylococcus*, scarlet fever, Rift Valley fever, and arboviruses, but none of these possibilities has gained many adherents among students of ancient epidemics.^{18,19} A number of these suggestions are diseases that may not have existed in the ancient world but appeared only in the 20th century. Lassa fever (1965) and Ebola fever (1976) are viral hemorrhagic fevers caused by RNA viruses. Rift Valley fever (1915) is a viral zoonosis whose vector is the bite of infected mosquitoes, typically the *Aedes* or *Culex* genera. Tularemia (1911) is bacterial with ticks as a vector. Each of these diseases has problems either in its symptomatology or in its epidemiology, as we will see later. For a full discussion of the various suggestions, see Morens and Littman,¹⁹ Poole and Holladay,²⁰ Cunha,²¹ Sallares,²² and Bruce-Chwatt and de Zulueta.²⁴

Beginning in the second half of the 20th century, new methodologies were developed that allowed fresh examinations of diseases, plagues, and pandemics in the ancient world. These methodologies were based on new academic disciplines, including forensic anthropology, demography, and epidemiology. Recent studies of the Plague of Athens reflect these developments in methodology. Although the emergence of modern medicine in the 19th century and the beginnings of the discipline of epidemiology led to a renewed interest in the Plague of Athens, it was not until 1992 that the first epidemiological approach to the disease was brought to bear by Dr. David Morens, now at the National Institute of Health, and myself^{19,25} using such devices as mathematical modeling to look at various possibilities. Mathematical modeling allowed us to examine infection and attack rates based on the various candidates and to determine how long it takes a particular disease to spread in a city and how long it would be able to remain endemic. Several diseases, such as measles, had to be eliminated as possibilities because

the population of Athens was too small to sustain an epidemic of them for more than a few months. Thus, the best possibilities were narrowed to typhus, typhus-like diseases, and smallpox.

EPIDEMIOLOGICAL INTERPRETATION

Person

A number of prominent people contracted the Plague of Athens. In the first outbreak, the historian Thucydides, who wrote a history of this period entitled *History of the Peloponnesian War*, contracted the disease. He survived and lived to write an account of it. The great Athenian statesman Pericles was less fortunate. He died in the first outbreak at the age of approximately 65, as did his sister, who was around 60, and 2 of his children, who were around 30 (Xanthippus) and 25 (Paralos; see Plutarch, *Life of Pericles*, 36.3-4).²³ The father of medicine, Hippocrates, according to legend, tried to combat the plague but in the end fled Athens so as not to contract it.²⁶ Thucydides (2:51) states that the epidemic "carried away all alike." Apparently, it spared no segment of the population. Thucydides relates that physicians had increased risk. The city of Athens was under siege, and the surrounding countryside was either burned or seized by the Spartans. Thucydides does not comment on the diet, although even with an unblockaded port, probably there were shortages of fruits and vegetables, with a possible concomitant deficiency of ascorbic acid (vitamin C). Vitamin A would have been obtained through fish and fish products.

Place

Thucydides (2.48.1) relates that the plague originated in Ethiopia and spread to Egypt and Libya and thence to the territory of the king of Persia (all of the Middle East and as far east as the Indus river). From there, it reached Athens, which was the hardest hit. The Roman historian Livy (*History of Rome*, 4.20-21, 4.25.3-4, and 4.30.8-10), writing many centuries later, reported that epidemics occurred in Rome in 433 and 438 BC. These outbreaks may well have been part of the same epidemic. The disease struck first at Athens' port, the Piraeus, one of the major ports of the Mediterranean. Premodern epidemics and pandemics of many diseases (cholera, dengue, plague, and smallpox) typically were spread by ships. Around July 430 BC, at the height of the epidemic, the Athenian general Hagnon took a naval expedition and sailed northward to Thrace with 4000 troops. The expedition was under sail for about 5 days, and by the time it had reached the city of Potidea, the

epidemic had struck. It killed about 25% of the army and infected the Athenian besiegers of Potidea. The city of Potidea itself, sealed shut by the invaders, was not afflicted by the epidemic.

Thucydides describes the directionality of spread when the disease first attacked Athens. It began in the port and spread inland into the area between the port and the city, which was encompassed by a defensive wall, and into the city itself. Thucydides is silent on the incidence rate (ie, number of new cases per population at risk per time), the high case fatality rate (proportion of those who died), and the rapidity of epidemic progression. He does say that the spread among the refugees in the city was caused by "poor ventilation," which is consistent with either airborne spread, poor hygiene and sanitation, crowding, or a combination of these. The water supply was probably uncontaminated. Potable water in the Piraeus came from cisterns and in Athens came from wells throughout the city. The city in addition had river-fed aqueducts from the Ilissus river and several springs. Thucydides mentions that despite frequent contact with the Spartan army, the disease did not spread much beyond the city. Thucydides (2.54.5) says that it did not enter the Peloponnesus to any extent and that its "full force was felt at Athens, and, after Athens, in the most densely populated of the other places." Perhaps the fact that the Peloponnesus was rather sparsely settled contributed to a lower epidemic rate.

Time

The epidemic broke out soon after the Spartans besieged Athens in early May 430 BC. There was another wave in the summer of 428 BC and in the winter of 427-426 BC. Apparently, the epidemic lasted around 4.5 to 5 years. The population of Athens,^{27,28} with refugees from the invading Spartan army, amounted to around 300,00 to 400,000. According to Thucydides (3.87.3), the plague was uninterrupted, but with exposing outbreaks:

The plague broke out among the Athenians for a second time. In fact, it had never entirely stopped . . . This second outbreak lasted for no less than a year and the first outbreak for two years.

Other Epidemiological Information

Thucydides portrays a virgin soil epidemic, which Thucydides confirms with his own words. It had a high attack rate and an unvarying course in persons of different ages, sexes, and nationalities. We cannot determine the incubation period, attack rate, or case fatality rate. From July to August 430 BC,

Thucydides portrays a virgin soil epidemic . . . It had a high attack rate and an unvarying course in persons of different ages, sexes, and nationalities.

26% of an expedition of 4000 hoplites were killed in 40 days. Thucydides (3.87.3) records a final death toll of 4400 hoplites (34%) and 300 cavalry deaths (30%). This suggests that the first wave was the worst. Survivors did not contract the disease a second time. This implies an attack rate of more than 25 and a case fatality rate of greater than 25%. Thucydides (2.50.1-2) also suggests ambiguously that there might have been a zoonotic component:

Though there were many dead bodies lying around unburied, the birds and animals that eat human flesh either did not come near them or, if they did taste the flesh, died of it afterwards. Evidence for this may be found in the fact that there was a complete disappearance of all birds of prey: they were not to be seen either round the bodies or anywhere else. But dogs, being domestic animals, provided the best opportunity of observing this effect of the plague.

Thucydides did not claim that anyone saw an infected dog or bird, but some scholars have used this to suggest epizootic involvement. However, Thucydides' description is so precise that if he saw an infected animal, he surely would have indicated it.

Thucydides' Text

Although Thucydides proposed to make a description that would allow others to identify the disease if it should break out again, he was not successful, although perhaps he was by ancient standards^{17,29} Thucydides jumbled physical signs and symptoms, epidemiological observations, and historical facts. He was not a physician, and a technical medical vocabulary probably did not exist in the 5th century BC when he was writing.³⁰ He thus applied common words to signs and symptoms with the result that it is often difficult to figure out precisely what he meant. For example, Thucydides used the term *phlyktainai*, a word to describe blisters and spots on a loaf of bread. Scholars have translated this work with various terms, including blains, blebs, blisters, bullae, eruptions, pimples, pustules, vesicles, and wheelks. Hippocrates used the word to describe burns and contact dermatitis. Thus, questions of how to interpret the word have led scholars to develop theories suggesting smallpox, typhus, and syphilis. Trying to

reach a definitive conclusion from the clinical description alone has reached a limit of what it is able to achieve. Only when the clinical is melded with the epidemiological can more definitive conclusions be reached. Although the epidemiological approach cannot be definitive, it can clearly eliminate whole categories of disease and some specific diseases, such as measles.

Common Source Acquisition

Common source acquisition is not likely on the basis of 'Thucydides' description. Although ergotism has been suggested, it is difficult to see how all grain sources could have been contaminated.

Waterborne disease is also unlikely because there was no centralized water system that could have become contaminated. Also, the disease spread from the Piraeus uphill to Athens. There is no way for hundreds of Athenian wells to be simultaneously cross-contaminated. Also, common source acquisition of a single epidemic disease in 430-425 BC can be ruled out, including such diseases as cholera, dysentery, ergotism, shigellosis, scurvy, and typhoid fever. However, recent DNA investigations of skeletal remains from the plague years have suggested the presence of typhoid. We will return to a discussion of these results later.

Person-to-Person Transmission

Of the 3 main categories of person-to-person transmission, enteric and inoculation transmission can be ruled out. Most consistent with the symptoms and description of the disease is aerosol/respiratory transmission, as suggested by the high attack rate and rapid spread. However, most respiratory diseases rapidly die out in crowded populations, although some, such as tuberculosis, can become endemic. Thucydides records that the epidemic lasted 2 years or more. In addition, the besieging Spartan army was not infected to any degree.

Mathematical Modeling

Because epidemic diseases vary in incubation periods and in susceptible percentages of populations, the rate of disease spread varies. For example, if a disease takes 2 days to become contagious and becomes capable of transmitting the disease to a third party, it might spread faster than an infectious disease that takes 10 days to become contagious. Using mathematical models based on these factors, we can predict for any contagious disease mathematical patterns of occurrence over specified time periods for various

Using mathematical models based on these factors, we can predict for any contagious disease mathematical patterns of occurrence over specified time periods for various population sizes and degrees of crowding.

population sizes and degrees of crowding. We can then apply such theoretical patterns to the Athenian epidemic.

The equation used to generate the epidemic curves is as follows³¹:

$$C_{t+1} = S_t(2 - qC_t)$$

where C_t is the number of cases of the disease at time t (the beginning of the epidemic), C_{t+1} is the number of cases of the disease at time $t + 1$, t is the chosen time interval (here the serial generation time), S_t is the number of susceptible persons at time t , and $q = 1 - p$ is the probability of adequate contact (ie, adequate to cause infection) between any 2 individuals per time interval t .

For each disease, the following assumptions are made:

1. There is universal susceptibility at the outset with infection either killing or conferring permanent immunity.
2. The serial generation time is assumed to be 4.5 days for influenza, 12 days for smallpox, 14 days for measles, and 19 days for streptococcal infections such as scarlet fever.
3. The total population of Athens was at least 100,000, and most likely during the war years, the population climbed to 200,000 to 400,000 as refugees poured into the city.
4. Adequate contact numbers are based on conservative estimates of the frequency of contact under conditions of known severe crowding.

The area of the Piraeus, connected by walls to the city of Athens, was about 4 square miles; this means a population density of 25,000 to 50,000 per square mile. Thirty percent of the square miles of Athens proper, not including Attica, was probably unoccupied.³² This reflects a typical settlement pattern of cities with crowded populations; people are crammed into houses, rooms, and public places, while at the same time there are large pockets of unoccupied spaces. There were probably about 10,000 dwellings, each occupied by 10 to 40 persons during the siege.³³ Although the main outbreak of

the plague may give some ambiguity about onset, we have a clearer case with the expedition of Hagnon. The expedition would not have set out from the Piraeus with infected people on board. Among the 4000 hoplites in the force and the 3000 hoplites left in Potidea, the epidemic curve from the index case to extinction was about 6 weeks. The results of applying the aforementioned formula can be seen in Figures 1 to 3.

Discussion

Epidemiological analysis is consistent with (1) an animal or insect reservoir or (2) respiratory transmission combined with a reservoir-like mechanism of persistence. This is consistent with the fact that the besieging army of Spartans was not infected.

Respiratory Transmission

Diseases transmitted only by a respiratory route would not have "over-wintered" and would have spread rapidly through the population and lasted over 2 years. Within a very short period, all the population would have been exposed. For example, in 1918, influenza spread through and died out in Newark, NJ (population of 435,000) in 12 weeks.

Measles could not have persisted for 2 or more years: 300,000 to 500,000 people are considered to be necessary to sustain measles transmission. Measles would have been extinguished in several months.

Smallpox remains a strong possibility. In the mathematical modeling (see Figures 1–3), a 12-day serial generation time is assumed as well as 2 to 20 adequate contacts and a minimal base population of 100,000 people per 4 square miles. Under this model, with 2 contacts, the epidemic would last 11 months. Smallpox outbreaks in virgin populations tend to be brief. For example, in Aztec Mexico in 1530, between 3.5 and 15 million out of a population of 25 to 30 million died within 6 months.⁵⁴ However, in the case of smallpox, unlike other respiratory diseases, viruses in dried smallpox secretions can survive for at least several months in clothes or bed linen.⁵⁵ There is even a case of smallpox being used for biological warfare against the Indians in the French and Indian War, when the commander of British forces gave the Indians blankets contaminated with smallpox.⁵⁶ It is possible in Athens that smallpox was carried back to the countryside and dispersed in an indolent transmission among the population of Attica in the remote areas in which the Spartan army was not present, although most of the population had come within the city walls.⁵⁷

Reservoir Transmission

The epidemiology of the epidemic is consistent with reservoir transmission (insect or animal vector). Anthrax might be a possibility because by a combination of inhalation, inoculation, and ingestion, it could be consistent with the clinical picture described by Thucydides. However, no large-scale anthrax epidemics are known to have ever occurred. Also, there is no easily identified reservoir in ancient Athens. Although sheep and cattle are very susceptible to anthrax, these animals were sent out of Attica to Euboea. Dogs, rodents, and birds remained in the city, but these are generally resistant to anthrax, although vultures and flies can spread the disease after feasting on contaminated carcasses, as is the case with dogs. Infected birds could contaminate the water supply, but Thucydides mentions the absence of birds. Anthrax is difficult to reconcile with these conditions.

Other reservoir diseases of epidemic proportion include various insect-borne diseases and diseases spread by various zoonotic reservoir hosts, such as malaria, plague, typhus, and various arboviral diseases (eg, dengue, yellow fever, and Rift Valley fever). All except Rift Valley fever are linked to war, refugees, and overcrowding. Dengue and Rift Valley fever both have explosive behavior consistent with the Plague of Athens. However, the clinical symptoms of both dengue and Rift Valley fever seem to differ greatly from Thucydides' description, especially in the nature of the rash.

Malaria is unlikely as a cause, particularly because it was not an unknown disease.⁵⁸ Although the ancient world did not distinguish between typhus and typhoid, it is clear that it was prevalent in ancient Greece and Rome. Hippocrates in the 5th century BC describes at least 6 cases in *Epidemics* I and II.⁵⁹

Of the reservoir diseases, typhus is the best fit. It is classically associated with war and crowding, and some symptoms described in Thucydides are suggestive, such as gangrene (if in fact Thucydides is describing gangrene) and blindness. Also, the disease is persistent in crowded populations for prolonged periods. The main argument against typhus is the presence of bullae in the Athenian epidemic. Although not regular, they are seen occasionally in typhus,¹⁵ and their occurrence in a mild form of the disease, rickettsial pox, might suggest that an ancestral rickettsial agent could have caused the Plague of Athens.

In summary, the epidemiological argument excludes all common source diseases and most respiratory diseases. By a process of elimination, the Plague of Athens can be limited to either a

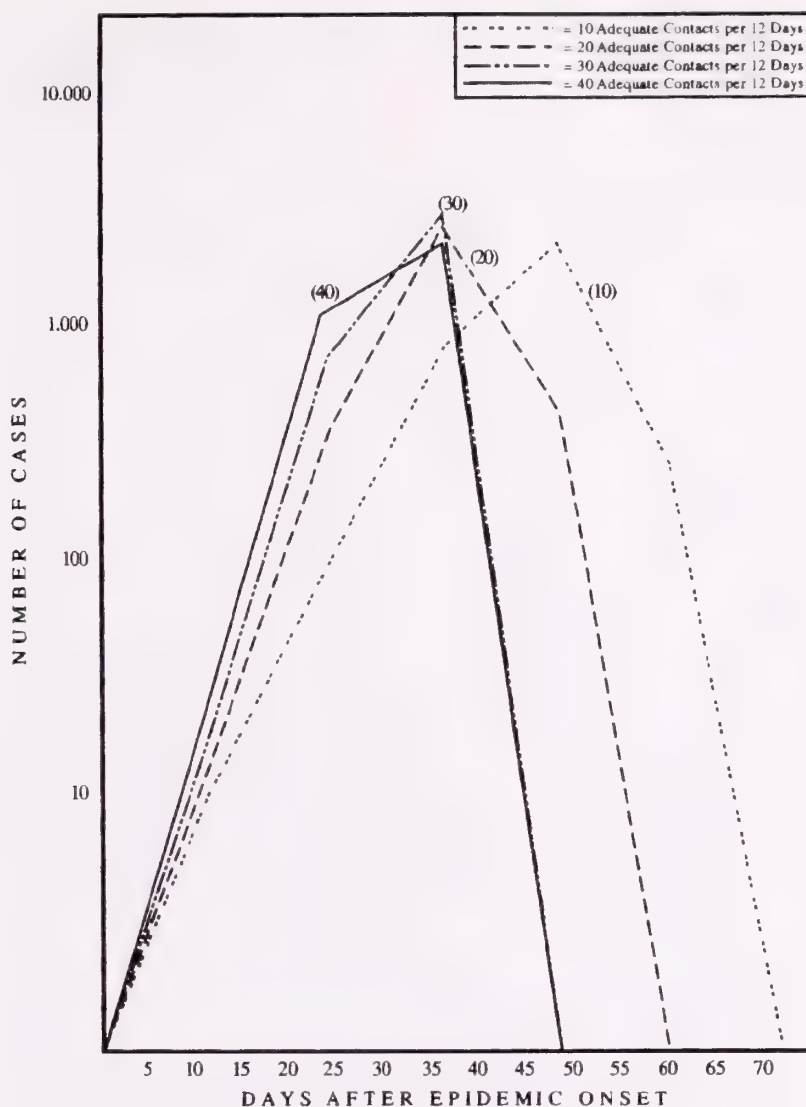


Fig 1. Theoretical epidemic curves of a transmissible respiratory disease with a 12-day serial generation time (eg, smallpox) in Hagnon's naval expedition to Potidea in 430 BC (a closed population of 4000 persons) according to the deterministic mathematical model of Maia. Four separate curves have been computed with adequate contact numbers of 10, 20, 30, and 40 per 12 days.

By a process of elimination, the Plague of Athens can be limited to either a reservoir diseases (zoonotic or vector-borne) or one of the respiratory diseases associated with an unusual means of persistence, either environmental/fomite persistence or adaptation to indolent transmission among dispersed rural

populations. The first category includes typhus, arboviral diseases, and plague, and the second category includes smallpox. Both measles and explosive streptococcal disease appear to be much less likely candidates.

reservoir diseases (zoonotic or vector-borne) or one of the respiratory diseases associated with an unusual means of persistence, either environmental/fomite

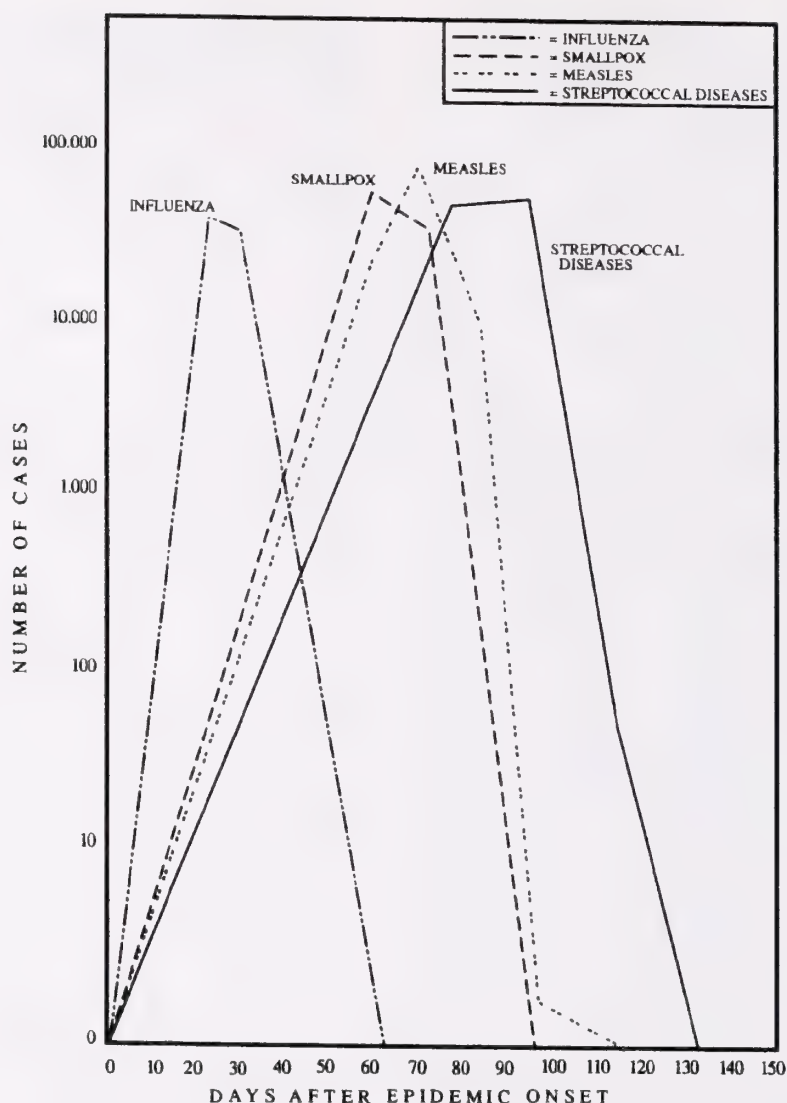


Fig 2. Theoretical epidemic curves of influenza A, smallpox, measles, and streptococcal disease in a closed population of 100,000 susceptible persons according to the mathematical model of Maia. The curves assume an adequate contact number of 10.

persistence or adaptation to indolent transmission among dispersed rural populations. The first category includes typhus, arboviral diseases, and plague, and the second category includes smallpox. Both measles and explosive streptococcal disease appear to be much less likely candidates.

PALEOPATHOLOGY

A challenge in historical epidemiology and the history of medicine is working with written history with little or no physical evidence. In some cases, particularly ancient Egypt, mummies provide actual physical evidence, as do skeletal remains. The new science

of DNA analysis has opened new windows in the investigation of the past.

Beginning in the second half of the 20th century, new methodologies were developed that allowed fresh examinations of diseases in history. These methodologies were based on new academic disciplines, including forensic anthropology, demography, and epidemiology. Scientific discoveries and technology allowed us new insights into the very essence of life. Chief among these were the discovery of DNA and the ability to analyze it, magnetic resonance imaging, and computed axial tomography (ie, three-dimensional pictures of hard and soft tissue inside the human body). These disciplines, discoveries, and technology have been brought to bear on the

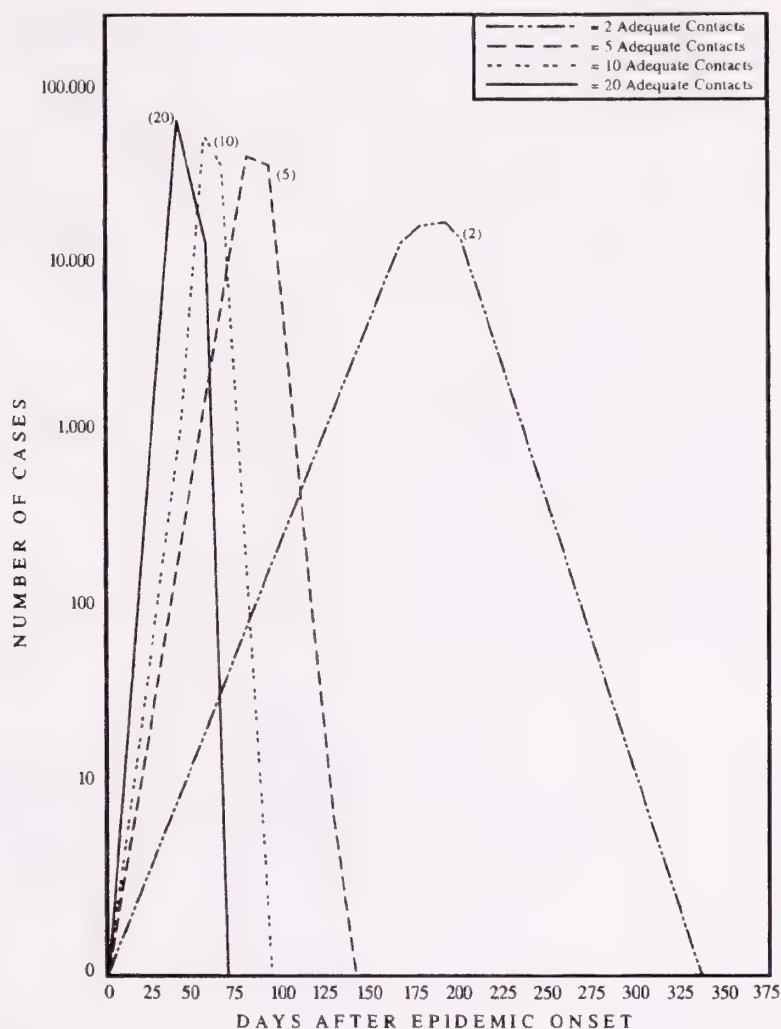


Fig 3. Theoretical epidemic curves of smallpox in a closed population of 100,000 susceptible persons according to the mathematical model of Maia. The curves assume adequate contact numbers of 2, 5, 10, and 20 persons per serial generation time of 12 days.

study of diseases in antiquity, including the nature and impact of plagues and pandemics in the ancient world.

The study of ancient DNA has begun to provide new clues in studying plagues of antiquity. Ancient microbial DNA has been successfully extracted⁴⁰; 400-year-old DNA of *Yersinia pestis*, the microorganism that causes bubonic plague, has been extracted from dental pulp.⁴¹ The extraction of human DNA is more problematic. There are claims that Neanderthal DNA has been extracted, and other claims that it is not possible to extract DNA more than 800 years old. In any event, for both the Plague of Athens and other ancient diseases, microbial DNA might provide a useful tool.

In 2001, a mass grave was discovered at the cemetery of the Kerameikos, which belonged to

the plague years. The excavation was conducted by Effie Baziotopoulou-Valavani,⁴² a member of the Greek Archaeological Service. The study of the skeletal material was undertaken by Professor Manolis Papagrigorakis⁴³ of the University of Athens. He was able to extract ancient microbial typhoid (*S. enterica* serovar Typhi) DNA from the remains. He posited that the probable cause of the Plague of Athens was this disease. It is premature, however, to draw this conclusion. We know from Hippocrates that typhoid was most likely endemic in the Greek world. The presence of an endemic disease does not necessarily indicate that it was the cause of death. A second problem is that the sampling was extremely small: 3 skeletons. A third difficulty is that the analysis of Papagrigorakis has been challenged. Shapiro *et al.*⁴⁴ argued that Papagrigorakis reported a 7%

divergence between their sequences and *S. enterica* serovar Typhi and therefore by simple phylogenetic analysis the ancient sequence falls outside both *S. enterica* and *S. typhimurium*. Therefore, the obtained sequences do not allow an identification of typhoid in these skeletons, although it is possibly a related *Salmonella*. Shapiro *et al.* further suggested that the sequence might have come from soil surrounding the burial and that the analysis is therefore compromised. Papagrigorakis *et al.*⁴⁵ responded that soil wash was used as a negative control and that there was no contamination, a phylogenetic analysis of only 1 gene might be insufficient or misleading, and *Salmonella* species do not survive for long in soil. A DNA analysis of the skeletal materials is an intriguing start that one could anticipate will lead to new information about paleopathology and the plague years. Because the plague struck Egypt according to Thucydides, what we most hope for are Egyptian mummified remains of someone who died of the Plague of Athens.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Bagnall R, Frier BW. *The Demography of Roman Egypt*. Cambridge, England: Cambridge University Press; 2008.; *Cambridge Studies in Population, Economy and Society in Past Time*.
- Littman RJ, Littman ML. Galen and the Antonine Plague. *Am J Philol* 1973; 94: 243–255.
- Nutton V. *Ancient Medicine*. London, England: Routledge; 2004.
- Nutton V, ed. *Pestilential Complexities: Understanding Medieval Plague*. London, England: Wellcome Trust Centre for the History of Medicine at UCL 2008.
- Hatcher J. *Plague, Population, and the English Economy 1348–1530*. Basingstoke, England: Macmillan; 1977.
- Hippocrates (W. H. S. Jones trans.) *Hippocrates: Nature of Man IV (Loeb Classical Library)*. Boston: Harvard University Press; 1931; 27–29.
- Hippocrates (W.D. Smith trans.) *Hippocrates: Epidemics Book 2, 4–7 (Loeb Classical Library)* Boston: Harvard University Press; 1994.
- Hippocrates (W.H.S. Jones trans.) *Hippocrates Ancient Medicine - Airs, Waters and Places - Epidemics I and III* Boston: Harvard University Press; 1984.
- Ferngren GB. *Medicine & Health Care in Early Christianity*. Baltimore, MD: Johns Hopkins University Press; 2009; 13–16.
- Thucydides. *Thucydides: the History of the Peloponnesian War*. Vols I and II. CF Smith, trans. Boston, MA: Harvard University Press; 1919 and 1921; *Loeb Classical Library*.
- Rechenaur G. *Thukydides und die Hippokratische Medizin*. Hildesheim, Germany: Georg Ohms; 1991.
- Weidauer K. *Thukydides und die Hippokratischen Schriften*. Heidelberg, Germany: Carl Winter; 1954.
- Littman RJ, Littman ML. The Athenian plague: smallpox. *Trans Am Philol Assoc* 1969; 100: 261–275.
- Zinsser H. *Rats, Lice and History*. Boston, MA: Little, Brown and Company; 1934.
- MacArthur WP. The Athenian plague: a medical note. *Classical Q* 1954; 4: 171–174.
- Scarborough J. Thucydides, Greek medicine and the plague at Athens: a summary of possibilities. *Episteme* 1970; 4: 77–90.
- Morgan TE. Plague or poetry? Thucydides on the epidemic at Athens. *Trans Am Philol Assoc* 1994; 124: 197–209.
- Longrigg J. The great plague at Athens. *Hist Sci* 1980; 18: 209–225.
- Morens DM, Littman RJ. Thucydides' syndrome reconsidered: new thoughts on the 'Plague of Athens.' *Am J Epidemiol* 1994; 140: 621–28.
- Poole JCF, Holladay AJ. Thucydides and the Plague of Athens. *Classical Q* 1979; 29: 282–300.
- Cunha BA. The cause of the Plague of Athens: plague, typhoid, typhus, smallpox, or measles? *Infect Dis Clin North Am* 2004; 18: 29–43.
- Sallares R. *The Ecology of the Ancient Greek World*. Ithaca, NY: Cornell University Press; 1991.
- Plutarch (B. Perrin trans.) *Plutarch: Parallel Lives, III: (Loeb Classical Library)*. Cambridge, Massachusetts: Harvard University Press; 1995.
- Bruce-Chwatt LJ, de Zulueta J. *The Rise and Fall of Malaria in Europe: A Historico-Epidemiological Study*. Oxford, England: Oxford University Press; 1980.
- Morens DM, Littman RJ. Epidemiology of the Plague at Athens. *Trans Am Philol Assoc* 1992; 122: 271–304.
- Pinault JR. *Hippocratic Lives and Legends*. Leiden, The Netherlands: Brill; 1992; chap 2.
- Hansen MH. Athenian population losses 431–403 B.C. and the number of Athenian citizens in 431 B.C. In: Hansen MH, ed. *Three Studies in Athenian Demography*. Copenhagen, Denmark: Munksgaard; 1988; *Historisk-Filosofiske Meddelelser*; vol LVI.
- Gomme AW. *The Population of Athens in the Fifth and Fourth Centuries BC*. Oxford, England: Basil Blackwell; 1933; 22–44.
- Jouanna J. *Hippocrates*. DeBevoise MB, trans. Baltimore, MD: Johns Hopkins University Press; 1999; 207–209.
- Dumortier J. *Le Vocabulaire Medical d'Eschyle et les Épigrammes critiques Hippocratiques*. Paris, France: Les Belles Lettres; 1975.
- Costa Maia JDO. Some mathematical developments on the epidemic theory formulated by Reed and Frost. *Hum Biol* 1952; 24: 167–200.
- Webster TBL. *Athenian Culture and Society*. Berkeley, CA: University of California Press; 1973.
- Siegel E. Epidemics and infectious diseases at the time of Hippocrates. Their relation to modern accounts. *Gesnerus* 1960; 17: 77–98.
- Magner JA. *Men of Mexico*. Milwaukee, WI: Bruce Publishing; 1942.
- Young M. Small pox conveyed by raw cotton. *Br Med J* 1951; 25: 324–341.

36. Duffy J. Smallpox and the Indians in the American colonies. *Bull Hist Med* 1951; 25: 324–341.
37. Hopkins DR. *Princes and Peasants: Smallpox in History*. Chicago, IL: University of Chicago Press; 1983; 234–294.
38. Sallares R. *Malaria and Greek History*. Manchester, England: Manchester University Press; 2002.
39. Cunha BA. Osler on typhoid fever: differentiating typhoid from typhus and malaria. *Infect Dis Clin North Am* 2004; 18: 111–125.
40. Zink AR, Sola C, Reischl U, et al. Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* 2003; 41: 359–367.
41. Drancourt M, Aboudharam G, Signoli M, et al. Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. *Proc Natl Acad Sci U S A* 1998; 95: 12637–12640.
42. Baziotopoulou-Valavani E. A mass burial from the cemetery of Kerameikos. In Stamatopoulou M, Yeroulanou M, eds. *Excavating Classical Culture: Recent Archaeological Discoveries in Greece*. Oxford, England: Archaeopress; 2002; 187–201. *Studies in Classical Archaeology*; vol I.
43. Papagrigorakis MJ, Yapijakis C, Synodinos PN, Baziotopoulou-Valavani E. DNA examination of ancient dental pulp indicates typhoid fever as a probable cause of the Plague of Athens. *Int J Infect Dis* 2006; 10: 206–214.
44. Shapiro B, Rambaut A, Gilbert TP. No proof that typhoid caused the Plague of Athens. *Int J Infect Dis* 2006; 10: 334–335.
45. Papagrigorakis MJ, Yapijakis C, Synodinos PN, Baziotopoulou-Valavani E. Insufficient phylogenetic analysis may not exclude candidacy of typhoid fever as a probable cause of the Plague of Athens. *Int J Infect Dis* 2006; 10: 335–336.

Historical Perspectives on Malaria: The Rockefeller Antimalaria Strategy in the 20th Century

Darwin H. Stapleton, PhD

Emeritus, The Rockefeller University, New York, NY

OUTLINE

INSUSCEPTIBILITY OF MALARIA TO CONTROL
BY INOCULATION

FOCUS ON THE MOSQUITO VECTOR

**Strategy of the Rockefeller Foundation
Typhus and the Advent of
Dichlorodiphenyltrichloroethane
Application of
Dichlorodiphenyltrichloroethane to
the Control of Malaria**

DECLINE IN THE USE OF

DICHLORODIPHENYLTRICHLOROETHANE: ATTENTION TO
NEW APPROACHES

Key Words: dichlorodiphenyltrichloroethane, malaria, Rockefeller, Foundation, typhus.

Malaria, a disease transmitted from human to human by mosquitoes, is currently estimated to be responsible for 350 to 400 million infections annually across the globe, and at least 1 million of those infections result in death for the infected individuals.¹ Malaria is a major and continuing challenge for medicine in the 21st century.

Malaria was definitely identified as a specific disease in the latter 1800s and became the object of coordinated attempts to control or eliminate it in the 20th century. However, it remains endemic in tropical regions across the globe and is intertwined with socioeconomic conditions, affecting particularly the poverty-stricken and those with little or no access to health care. The history of malaria control for the last 100 years is suggestive of the complex set of problems associated with zoonotic diseases.^{2,3} This historical review focuses on the Rockefeller Foundation's antimalaria strategy in the first half of the 20th century, a strategy that has deeply influenced subsequent global attempts at malaria control.

Until the mid-1800s, malaria was relatively inseparable from various other fevers, particularly epidemic fevers that struck in the summers in both tropical and temperate zones. For example, the famous yellow fever epidemic that in 1793 struck Philadelphia, then the capital of the new United States, and killed about 10% of the population, probably had both malaria and yellow fever victims, although the deaths of the latter were far more dramatic and more frequent.^{4,5} Contemporary misunderstanding of the etiology of the disease made effective short-term responses impossible: it took a quarter-century before the city's investment in a municipal water system and improvements in sewerage (which drastically reduced breeding places for mosquitoes) proved to be the means of ending

ABSTRACT

Malaria, a serious disease for all of human history, was not effectively handled until methods for control of its insect vector, the *Anopheles* mosquito, were developed at the beginning of the 20th century. The Rockefeller Foundation's antimalaria program spread vector-control strategies throughout the world, and its adoption of dichlorodiphenyltrichloroethane during World War II created an especially powerful and effective malaria control strategy. However, insect resistance to dichlorodiphenyltrichloroethane and restrictions on dichlorodiphenyltrichloroethane use due to its long-term environmental effects are factors in the persistence of malaria as a serious health problem. *Mt Sinai J Med* 76:468–473, 2009. © 2009 Mount Sinai School of Medicine

Address Correspondence to:

Darwin H. Stapleton
Rockefeller Archive Center
Sleepy Hollow, NY
Email: stapled@rockefeller.edu

the epidemics.^{6,7} Political and technological changes, rather than increased knowledge of disease, ended the fever years in Philadelphia.

The potential foundation for disease control developed in the mid-19th century with the fruitful intersection of scientific and investigative medicine. As the germ theory of disease origin and transmission took hold, new possibilities for understanding and controlling disease emerged. Vector-carried diseases, such as malaria, were more complex matters than human-to-human diseases, but they were also susceptible to the new scientific-medical tools of analysis. Alphonse Laveran recognized the malaria parasite in the blood of infected persons in 1880. Giambattista Grassi did research that in 1898 recognized particular mosquito species as carriers of malaria, having established that they were carrying the malaria parasite. At virtually the same time, Ronald Ross demonstrated conclusively that there was a malaria pathway between humans and mosquitoes.⁸ The very fact that these milestones were accomplished by a Frenchman, an Italian, and a British subject working in India shows both the internationality of concern about malaria and that the tools of modern science could be brought to bear on disease research almost anywhere on the globe.

INSUSCEPTIBILITY OF MALARIA TO CONTROL BY INOCULATION

However, unlike several other major diseases of mankind that were attacked and vanquished within a few decades after Pasteur showed the way with a rabies vaccine, malaria did not prove susceptible to control by some inoculation.

The malaria parasite, having evolved with human beings over several thousand years, is a wily adversary that thus far has defeated all direct assaults of scientific medicine.

The malaria parasite, having evolved with human beings over several thousand years, is a wily adversary that thus far has defeated all direct assaults of scientific medicine.

Until the 1930s, only quinine, derived from the bark of the *Cinchona* tree of South America, whose curative properties were known in pre-Columbian times, provided a consistent drug-based means of

controlling the disease for those already infected.⁹ Early in the 20th century, the systematic distribution of quinine in some areas of the world allowed the malaria-infected to carry out life tasks without suffering malaria's debilitating symptoms.^{10,11}

FOCUS ON THE MOSQUITO VECTOR

The other approach to malaria control has been to attack the mosquito vector. Following the success of antimosquito campaigns in Cuba (1900–1901) and the Panama Canal Zone (1905–1914), which significantly reduced yellow fever incidence in those areas, primarily through mosquito control, antimalaria programs adopted the same environmental strategies, which ranged from screening windows to covering all water-collecting devices to draining swamps.^{12,13} Later additions to such environmental strategies included managing irrigation practices to interrupt mosquito larvae propagation and introducing the mosquito larvae-eating fish *Gambusia*. Most of these measures, which did nothing for those already infected but which attempted to limit or stop the transmission of malaria by reducing the mosquito population, required constant vigilance.

Strategy of the Rockefeller Foundation

In 1915, the International Health Board, a philanthropic institution funded by John D. Rockefeller, took on the challenge of malaria control, first by experimenting with various strategies in rural areas of the southeastern United States and then by extending its experiments to several malarious regions around the world.^{14,15} (In 1928, the International Health Board, having worked in tandem with the Rockefeller Foundation from the beginning, was merged with the Rockefeller Foundation. The foundation continued the board's antimalaria program for another quarter-century.)

Typical of the Rockefeller-funded projects was the work undertaken from 1924 to 1938 in Italy, where malaria had been a problem in agricultural and coastal regions since time immemorial. The Italian government had instituted a vigorous program of quinine distribution at the beginning of the century and also carried out long-term engineering projects to fill swamps and estuaries. However, the Rockefeller officer in charge, Lewis Hackett, belittled these ameliorative measures and initiated demonstration projects that focused on the mosquito vector. After identifying the malaria-carrying mosquitoes specific to Italy and after studying the favored environments

of those mosquitoes, Hackett centered his work on spraying mosquito habitats with Paris Green, a highly toxic double salt of copper and arsenic that had been identified recently as a mosquito larva killer.^{16,17} Hackett elaborated on his work in *Malaria in Europe*, which became virtually a handbook for malariologists of his era.^{18,19}

This strategy was derided by a 1930 League of Nations report, which argued that general sanitation measures and the distribution of quinine improved the health of more people than antimosquito campaigns did. Rockefeller officers countered by arguing that those approaches did little or nothing to break the malaria transmission cycle and pointed out that, in Italy and in their work in areas as different as India and Mexico, they had arguably successful antimosquito projects.^{20,21} The foundation remained wedded to mosquito control as its method of attacking malaria, in particular concentrating on the effective use of insecticides.

The antimalaria program was carried out in the framework of a global Rockefeller public health program that emphasized the control of several other major diseases, including yellow fever, ancylostomiasis (hookworm), and influenza.²² The program generally attempted to create model projects wherever the Rockefeller approach was welcomed and then to train a generation of public health leaders who would continue the scientific-technical strategy it favored. Malaria projects as widespread as the southern United States, Columbia, Italy, India, and the Philippines were expected to influence the course of public health work throughout the world.^{23–25}

Typhus and the Advent of Dichlorodiphenyltrichloroethane

Nonetheless, by the end of the 1930s, global malaria control remained a vision that was far from realization. It was the advent of dichlorodiphenyltrichloroethane (DDT)—a powerful new insecticide—during World War II that created unanticipated possibilities. The story of DDT begins with the threat of another zoonotic disease: typhus.

It was the advent of dichlorodiphenyltrichloroethane (DDT)—a powerful new insecticide—during World War II that created unanticipated possibilities.

At the outbreak of World War II, public health officials had every reason to anticipate that

major typhus epidemics would occur, particularly in European populations weakened by years of economic privation and by the opportunities for infection in crowded refugee camps, internment camps, prisoner-of-war camps, ghettos, slave-labor facilities, and concentration camps.²⁶ The best means of controlling typhus epidemics at that time remained strict personal sanitation; this was based on knowledge that typhus is transmitted by the human body louse, which has evolved with humans to favor woven fabric as a habitation. The alternatives to control by sanitation at the time were the use of insecticides and vaccines, but the insecticides most effective against lice were in general highly toxic to humans, and the vaccines available were still in trial or untested in potential epidemic circumstances.

The Rockefeller Foundation began a modest antityphus program in 1940 by connecting with the Harvard University laboratory of Hans Zinsser, who had for many years been investigating the family of *Rickettsia* microbes and had already published his classic historical work, *Rats, Lice and History*.²⁷ Over the next months, with the increasing likelihood of the entry of the United States into the war, typhus research grew into a significant element of the Rockefeller Foundation's public health program. Because of the foundation's extensive connections, it could survey, study, and otherwise learn about typhus virtually throughout the world. Moreover, it could test typhus vaccines in the field with collaboration from local authorities who already knew and trusted the motives of the foundation. These tests initially included promising but ultimately disappointing trials of existing vaccines in venues as disparate as vulnerable populations in Spain and young men at the New York State Vocational Institution in the Hudson River Valley town of West Coxsackie. By the time of the entry of the United States into the war in December 1941, the foundation had decided that an attempt to develop a typhus vaccination was unlikely to succeed in the short run. That decision soon was reinforced by a request from the US National Research Council Subcommittee on Tropical Diseases to investigate "the advisability of undertaking a study of various delousing methods in N[ew] Y[ork]."²⁸ The next month, the foundation created a laboratory for the study of louse-borne typhus as a unit of the foundation's research laboratory located on the campus of the Rockefeller Institute in New York City.²⁹

The foundation's antityphus laboratory, which soon became known as the louse lab, quickly established a collaborative agreement with the Bureau of Entomology and Quarantine of the US Department of Agriculture by which the bureau would test

insecticides and the foundation would test them and other louse-control methods on human subjects in field conditions. It quickly became apparent that the foundation's 30 years of fieldwork in public health and its global network of contacts were advantages that a government agricultural laboratory did not possess. After successfully establishing a colony of lice in New York, the foundation was able carry out tests of louse-control methods on conscientious objectors in a forestry camp in the United States and in a village in Mexico.

However, the foundation's antityphus program soon was radically altered by the identification of DDT as a powerful insecticide that could control the human body louse. In August 1942, J.R. Geigy Co. of Switzerland shipped to its office in New York a sample of this new insecticide, which then was sent to the Bureau of Entomology and Quarantine. Early in 1943, the foundation learned from the Bureau about DDT's wonderful properties, and it was able to carry out its own tests in April. By July 1943, a team of foundation officers in Algeria was testing DDT as a lousicide.

The result of the Rockefeller team's work was the creation of an effective insecticide-delivery methodology based on DDT that could control the human body louse. Their methodology was put to the test during an incipient typhus epidemic in Naples soon after the Allied invasion of Italy. A combination of wartime privation and overcrowding of the civilian population, a depleted local health system, and refugees from the Balkans, who presumably were the locus of the infection, created the conditions for a rapid rise in typhus cases in Naples in December 1943.³⁰ The Rockefeller team from North Africa was called in and "formed the backbone" of the intensive project that eventually dusted 1.3 million people with insecticides.³¹ This event has generally been recognized by historians as the first clear demonstration of the control of human disease by DDT.³²⁻³⁴

Application of Dichlorodiphenyltrichloroethane to the Control of Malaria

Yet it was also essentially the end of the foundation's antityphus experiments because DDT had convincingly demonstrated its effectiveness and further experimentation seemed unwarranted. The Rockefeller Foundation instead turned its attention to experiments with DDT for mosquito larva control in Italy, essentially renewing the antimalaria project that it had carried out there from 1924 to 1938. The foundation found DDT so effective that it decided

to attempt the eradication of mosquitoes from the entire island of Sardinia, a demonstration project that it carried out from 1946 to 1951.

In Sardinia, the Rockefeller Foundation, with financial support from the United Nations Relief and Reconstruction Administration and the Italian government, created a highly organized team that sought to locate and spray every breeding place of anopheline mosquitoes in Sardinia, with the goal of eradicating them from the island. In an area approximately the size of the state of New Hampshire, there were mountains, swamps, forests, villages, and small cities that had to be mapped, sampled, and efficiently sprayed. Visitors from throughout the world studied the administration, methodology, and technology associated with the Sardinian project, which was successful in eliminating the scourge of malaria from the island. After that great effort, the foundation yielded leadership in the malaria field to the recently created World Health Organization, whose global antimalaria programs, beginning in 1955, have continued the foundation's insecticide-based strategies.^{16,34,35} With only a glimmer of hope for a malaria vaccine, it appears that insecticides will be the mainstay of antimalaria work for the foreseeable future.^{36,37}

DECLINE IN THE USE OF DICHLORODIPHENYL- TRICHLOROETHANE: ATTENTION TO NEW APPROACHES

In the last decade, renewed recognition of malaria's terrible burden in tropical regions has brought both the US government and the Gates Foundation into the antimalaria fight, pouring billions of dollars into a range of approaches to malaria control.^{38,39} This has also raised the profile of DDT, whose use was declining, particularly after its deleterious environmental effects were recognized. In 1999, a group of scientists urged the United Nations to reconsider its plan to push for a global ban on DDT, and they were successful.⁴⁰

Malaria can no longer be regarded as a rural disease. . . . Acquired resistance by mosquitoes defeats each new generation of insecticides. . . . Many insecticides are persistent over years and migrate through ecological cycles to cause unanticipated effects.

Malaria obviously remains a major zoonotic disease that has been receiving renewed and highly appropriate attention in recent years. The pathway to serious inroads on the disease is by no means clear.

Among the major issues to be dealt with are the following:

1. Malaria can no longer be regarded as a rural disease; a case in point is Mumbai, India.⁴¹ In 1981, for the first time in years, anopheline mosquitoes, associated with rapid growth in construction projects, were found there early on with skyscraper projects that had been stopped before completion: monsoon rains turned concrete warrens into excellent mosquito breeding grounds. New construction begun in the 1990s seemed to support increased malaria incidence.
2. Acquired resistance by mosquitoes defeats each new generation of insecticides. Because of the rapid production of new generations, insects can develop immunity to insecticides rapidly: the first documented resistance to DDT occurred in 1945, only a few years after its introduction.
3. Many insecticides are persistent over years and migrate through ecological cycles to cause unanticipated effects. DDT is perhaps the most studied example: it is persistent in that it is itself long-lasting and in that its breakdown products, such as dichlorodiphenyldichloroethylene (also toxic), are also persistent through the food chain. Places intensely sprayed with DDT, as well as the sites of DDT production, are still hot 3 and 4 decades later. At least by the early 1970s, concerns were expressed publicly about the potential effects of DDT on human health.^{42,43} There is increasing evidence that DDT and its breakdown products are associated with certain cancers and birth defects when present at high levels.^{44,45} Nonetheless, it must be considered in any global plan for malaria control.⁴⁶
4. Malaria has acquired resistance to a series of antimalaria drugs, beginning with chloroquine and Atabrine in late 1930s. Low-level use can be a significant problem: poverty-stricken individuals may use just enough drugs to get through a crisis but not enough to clear the bloodstream of the malaria parasite, encouraging the development of resistance.

It is apparent that malaria as a disease and malaria as a controversial socioeconomic problem will remain with us indefinitely.

It is apparent that malaria as a disease and malaria as a controversial socioeconomic problem will remain with us indefinitely.⁴⁷

Recognizing the successes and failures of the last 100 years of its history will be an important element in future attempts to control or eradicate it.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Worldwatch Institute. *Vital Signs, 2007–2008: The Trends That Are Shaping Our Future*. New York, NY: W.W. Norton & Co.; 2007; 122.
2. Packard RM. *The Making of a Tropical Disease: A Short History of Malaria*. Baltimore, MD: Johns Hopkins University Press; 2007.
3. Webb JLA Jr. *Humanity's Burden: A Global History of Malaria*. Cambridge, England: Cambridge University Press; 2008.
4. Powell JH. *Bring Out Your Dead: The Great Plague of Yellow Fever in Philadelphia in 1793*. Philadelphia, PA: University of Pennsylvania Press; 1993.
5. Humphreys M. *Malaria: Poverty, Race, and Public Health in the United States*. Baltimore, MD: Johns Hopkins University Press; 2001; 23–29.
6. Blake NM. *Water for the Cities*. Syracuse, NY: Syracuse University Press; 1956.
7. Stapleton DH, ed. *The Engineering Drawings of Benjamin Henry Latrobe*. New Haven, CT: Yale University Press; 1980; 28–36.
8. Rosen G. *A History of Public Health*. Baltimore, MD: Johns Hopkins University Press; 1993; 298–300.
9. Slater LB. *War and Disease: Biomedical Research on Malaria in the Twentieth Century*. New Brunswick, NJ: Rutgers University Press; 2009; 59–83.
10. Poser CM, Bruyn GW. *An Illustrated History of Malaria*. New York, NY: Parthenon; 1999; 75–94.
11. Rocco F. *The Miraculous Fever-Tree: Malaria, Medicine and the Cure That Changed the World*. New York, NY: Harper Collins; 2003.
12. Crosby MC. *The American Plague: The Untold Story of Yellow Fever, the Epidemic That Shaped Our History*. New York, NY: Berkeley Books; 2006; 91–207.
13. McCullough D. *The Path Between the Seas: The Creation of the Panama Canal, 1870–1914*. New York, NY: Simon & Schuster; 1977; 405–426, 464–68.
14. Farley J. *To Cast Out Disease: A History of the International Health Division of the Rockefeller Foundation (1913–1951)*. New York, NY: Oxford University Press; 2004; 107–153.
15. Humphreys M. *Malaria: Poverty, Race, and Public Health in the United States*. Baltimore, MD: Johns Hopkins University Press; 2001; 74–76.
16. Stapleton D. Internationalism and nationalism: the Rockefeller Foundation, public health, and malaria in Italy, 1923–1951. *Parassitologia* 2000; 42: 127–134.
17. Donelli G, Serinaldi E. *Dalla Lotta alla Malaria alla Nascita Dell'Istituto di Sanità Pubblica: Il Ruolo della*

- Rockefeller Foundation in Italia: 1922–1934*. Rome, Italy: Editori Laterza; 2003.
18. Hackett LW. *Malaria in Europe: An Ecological Study*. London, England: Oxford University Press; 1937.
 19. Stapleton DH. Hackett, Lewis Wendell. In: Bynum WF, Bynum H, eds. *Dictionary of Medical Biography*. Westport, CT: Greenwood Press; 2007.
 20. Muraleedharan VR, Veeraraghavan D. Anti-malarial policy in the Madras presidency: an overview of the early decades of the twentieth century. *Med Hist* 1992; 36: 290–305.
 21. Birn A-E. Eradication, control, or neither? Hookworm vs. malaria strategies and Rockefeller public health in Mexico. *Parassitologia* 1998; 40: 137–147.
 22. Farley J. *To Cast Out Disease: A History of the International Health Division of the Rockefeller Foundation (1913–1951)*. New York, NY: Oxford University Press; 2004.
 23. Stapleton DH. Malaria eradication and the technological model: the Rockefeller Foundation and public health in East Asia. In: Yip K-C, ed. *Disease, Colonialism, and the State: Malaria in Modern East Asian History*. Hong Kong: Hong Kong University Press; 2009; 71–84.
 24. Rodríguez PM. De ratones, vacunas y hombres: el programa de fiebre amarilla de la Fundación Rockefeller en Colombia, 1932–1948. *Dynamis* 2004; 24: 119–155.
 25. Stapleton DH. Technology and malaria control: the career of Rockefeller Foundation engineer Frederick W. Knipe. *Parassitologia* 2000; 42: 59–68.
 26. A step in preparedness against typhus. *Rockefeller Found Trustees Confidential Bull* 1940; 16–18.
 27. Zinsser H. *Rats, Lice and History: A Study in Biography*. Boston, MA: Little, Brown, and Co.; 1934.
 28. Johannes H. Bauer diary, 22 January 1942, box 5, Record Group [hereafter RG] 12.1, Rockefeller Foundation Archives [hereafter RFA], Rockefeller Archive Center, Sleepy Hollow, New York, USA; extract from Trustees' Minutes, 11–12 September 1942, folder 261, box 32, series 100, RG 1.1, RFA.
 29. Stapleton DH. A lost chapter in the early history of DDT: the development of anti-typhus technologies by the Rockefeller Foundation's louse laboratory, 1942–1944. *Technol Cult* 2005; 46: 513–540.
 30. Soper FL, Davis WA, Markham FS, Riehl LA. Typhus fever in Italy, 1943–1945, and its control with louse powder. In: Kerr JA, ed. *Building the Health Bridge: Selections from the Works of Fred L. Soper, M.D.* Bloomington, IN: Indiana University Press; 1970; 384–415.
 31. Turner TB. Sicily and Italy. In: Taylor RR, Lada J, Hoof EC, eds. *Preventive Medicine in World War II*. Vol 8. Washington, DC: Office of the Surgeon General; 1976; 308.
 32. Russell EP III. 'Speaking of annihilation': mobilizing for war against human and insect enemies, 1914–1945. *J Am Hist* 1996; 82: 1523–1524.
 33. Williams TL. *A Short History of Twentieth-Century Technology, c. 1900–c. 1950*. New York, NY: Oxford University Press; 1982; 328.
 34. Farley J. Mosquitoes or malaria? Rockefeller campaigns in the American south and Sardinia. *Parassitologia* 1994; 36: 165–173.
 35. Brown PJ. Failure-as-success: multiple meanings of eradication in the Rockefeller Foundation Sardinia Project, 1946–1951. *Parassitologia* 1998; 40: 117–130.
 36. PATH Malaria Vaccine Initiative. Available at: <http://www.malariavaccine.org>. Accessed June 2009.
 37. European Malaria Vaccine Initiative. Available at: <http://www.emvi.org>. Accessed June 2009.
 38. Garrett L. The return of infectious disease. *Foreign Aff* 1996; 88: 66–79.
 39. Shell ER. Resurgence of a deadly disease. *Atl Mon* 1997; 280: 45–60.
 40. Open letter to DDT treaty negotiators. Available at: http://www.malaria.org/DDT_open.html. Accessed September 2008.
 41. Kamat V. Resurgence of malaria in Bombay (Mumbai) in the 1990s: a historical perspective. *Parassitologia* 2000; 42: 135–148.
 42. Commoner B. *The Closing Circle: Nature, Man and Technology*. New York, NY: Bantam Books; 1971; 163.
 43. Curtis J, Profeta T. *After Silent Spring: The Unsolved Problems of Pesticide Use in the United States*. New York, NY: Natural Resources Defense Council; 1993; 8–18.
 44. Eskenazi B, Chevrier J, Rosas LG, et al. The Pine River Statement: human health consequences of DDT use. *Environ Health Perspect*. doi://10.1289/ehp.11748. Available at: <http://www.ehponline.org>.
 45. Worldwatch Institute. *Vital Signs, 2007–2008: The Trends That Are Shaping Our Future*. New York, NY: W.W. Norton & Co.; 2007; 124–125.
 46. van den Berg H. Global status of DDT and its alternatives for use in vector control to prevent disease. *Environ Health Perspect*. doi://10.1289/ehp.0900785. Available at: <http://www.ehponline.org>.
 47. Okie S. A new attack on malaria. *N Engl J Med* 2008; 358: 2425–2428.

SPECIAL FEATURE

Paradigm Shift in the Pathophysiology of Postmenopausal and Thyrotoxic Osteoporosis

Mone Zaidi,¹ Jameel Iqbal,¹ Harry C. Blair,² Alberta Zallone,³ Terry Davies,¹ and Li Sun¹

¹Mount Sinai Bone Program, Mount Sinai School of Medicine, New York, NY

²Department of Pathology and Cell Biology, University of Pittsburgh, Pittsburgh, PA

³Department of Histology, University of Bari, Bari, Italy

OUTLINE

PATHOPHYSIOLOGY OF POSTMENOPAUSAL

OSTEOPOROSIS

Established Effects of Estrogen Loss on the Skeleton

Bone Loss During the Menopausal Transition

Stimulation of Bone Resorption by Follicle-Stimulating Hormone

Estrogen Effects Are Distinct from

Follicle-Stimulating Hormone Effects in Vivo Effect of Follicle-Stimulating Hormone on the Human Skeleton

PATHOPHYSIOLOGY OF THYROTOXIC

OSTEOPOROSIS

Hyperthyroidism and Osteoporosis in Humans

Pharmacological Evidence for Thyroid Hormone Action on Bone

Genetic Evidence for Thyroid Hormone Action on Bone

Thyroid-Stimulating Hormone Directly Suppresses Bone Remodeling: Genetic Evidence

Exogenously Administered Thyroid-Stimulating Hormone Suppresses Bone Resorption

Thyroid-Stimulating Hormone and Human Skeletal Homeostasis

Thyroid-Stimulating Hormone Receptor Polymorphisms and Bone Mass

PARADIGM SHIFT IN UNDERSTANDING

OSTEOPOROSIS

ABSTRACT

Bone loss is an inevitable result of getting older in both sexes. This loss is accelerated in women in the early years following menopause and in both sexes with drugs and diseases, prominently hyperthyroidism. Postmenopausal osteoporosis and hyperthyroid osteoporosis, both of which are widely recognized clinical entities, have been thought to arise from low estrogen levels and high thyroid hormone levels, respectively. However, when estrogen declines, follicle-stimulating hormone, an anterior pituitary hormone, rises. Likewise, when thyroid hormone levels rise in hyperthyroidism, thyroid-stimulating hormone levels fall. We have used a complement of mouse genetic, pharmacological, and cell biological approaches to unravel a direct action of follicle-stimulating hormone and thyroid-stimulating hormone on bone remodeling and bone mass. Apart from establishing a novel pituitary-bone axis of physiological significance, our studies challenge the existing 1-hormone, 1-disease paradigm in endocrinology and suggest that a plurality of hormonal disturbances involving pituitary hormones cause postmenopausal and hyperthyroid osteoporosis. New

Address Correspondence to:

Mone Zaidi

Mount Sinai Bone Program
Mount Sinai School of Medicine
New York, NY

Email: mone.zaidi@mssm.edu

therapeutic targets thus emerge from these studies. *Mt Sinai J Med* 76:474–483, 2009. © 2009 Mount Sinai School of Medicine

Key Words: bone loss, follicle-stimulating hormone, osteoclast, osteopenia, postmenopausal osteoporosis, thyroid-stimulating hormone, thyrotoxic osteoporosis.

One in 2 women over the age of 50 has osteoporosis, with a fracture incidence that far exceeds the combined incidence of breast cancer, heart disease, and stroke.¹ Osteoporosis thus constitutes a profound public health hazard, with at least 40 million Americans suffering from bone loss, 1.5 million new fractures per year, and a cost estimated at \$15 billion in direct medical care.

More than 10% of postmenopausal women receive thyroid hormone replacement therapy.² Furthermore, hyperthyroidism is 10-fold more common in women, with an incidence of 1 in 1000 women per year.² Overt hyperthyroidism and the therapeutic suppression of thyroid-stimulating hormone (TSH) are both associated with osteoporosis, which is more pronounced in menopause.^{3,4} Thus, postmenopausal osteoporosis and thyrotoxic osteoporosis overlap significantly in women of advancing age.

Over the past decade, our respective laboratories have focused on understanding the pathophysiology of these 2 important medical conditions, postmenopausal osteoporosis and hyperthyroid osteoporosis, and in doing so have discovered novel mechanisms of pathological bone loss.

The osteoporoses of menopause and hyperthyroidism, which have been thought solely to be due to declining serum estrogen and elevated serum thyroid hormones, respectively, can now be attributed in part to accompanying changes in 2 anterior pituitary hormones: follicle-stimulating hormone (FSH) and TSH.

The osteoporoses of menopause and hyperthyroidism, which have been thought to be solely due to declining serum estrogen and elevated serum thyroid hormones, respectively, can now be attributed in part to accompanying changes in 2 anterior pituitary hormones: follicle-stimulating hormone (FSH) and TSH. Here we focus on cell biological, mouse genetic, and

pharmacological studies that have established a direct action of the respective hormones on the skeleton, and we discuss emerging clinical evidence for a novel pituitary-bone axis in humans that bypasses master endocrine organs, namely the ovaries and thyroid gland.

PATHOPHYSIOLOGY OF POSTMENOPAUSAL OSTEOPOROSIS

We review studies that propose FSH is a physiological stimulator of osteoclastic bone resorption. We hypothesize that a rising FSH level contributes to the increased bone resorption and bone loss in hypogonadism, particularly during the menopausal transition, in which profound bone loss is accompanied by trabecular perforations and overt bone loss in the face of high FSH levels and apparently normal estrogen levels. Potential therapeutic implications include the development of antagonists to both circulating FSH and its osteoclastic receptor.

Established Effects of Estrogen Loss on the Skeleton

Since Fuller Albright's discovery of hypogonadal bone loss, estrogen deficiency has been considered its sole culprit.⁵ Evidence for over 2 decades has been overwhelming: reduced estrogen causes bone loss, and estrogen replacement prevents this loss.⁶ This concept holds well across species. In contrast to in vivo and clinical data, in vitro studies explaining how estrogen prevents bone loss are not convincing.

There is evidence for weak but direct inhibitory effects of estrogen on osteoclast precursors, which are exerted by inhibition of the phosphorylation of the enzyme Janus N-terminal kinase downstream of the receptor activator for nuclear factor kappa B ligand (RANK-L).^{7,8} By this mechanism, estrogen inhibits osteoclastogenesis, which increases dramatically upon hormone withdrawal. Estrogen also uses multiple pathways to alter the secretion of RANK-L and its decoy receptor osteoprotegerin from osteoblasts,⁹ thereby secondarily affecting osteoclast formation. Effects on osteoclastogenesis are also exerted via T cells, in which estrogen has been shown to regulate the production of the cytokine tumor necrosis factor α (TNF α) and also to modulate the production of nitric oxide.^{10–12} Likewise, an anabolic effect of estrogen is well established in animal models.¹³ We neither dispute the anabolic actions of estrogen nor doubt its antiresorptive function.

Bone Loss during the Menopausal Transition

Sex steroid withdrawal in mammals, either acutely through gonadal ablation or gradually as in natural menopause, is invariably associated with a sharp rise in the serum levels of FSH. A primary function of FSH, hitherto considered its only role, is to stimulate ovarian folliculogenesis and estrogen production. This results in a positive feedback regulation of FSH secretion from the pituitary in response to failing ovaries. Thus, in women undergoing menopause, serum FSH levels rise almost exponentially in response to failing ovaries; this occurs, however, before serum estrogen levels begin to decline.

It has been determined recently that, during the menopausal transition, despite no reductions in circulating estrogen and in the presence of elevated FSH levels, there is a profound increase in the rate of bone loss, most prominently during late perimenopause.^{14,15}

It has been determined recently that, during the menopausal transition, despite no reductions in circulating estrogen and in the presence of elevated FSH levels, there is a profound increase in the rate of bone loss, most prominently during late perimenopause.

In fact, up to 50% of lifetime bone loss can occur within the first 5 years of menopause. Parameters of bone remodeling, such as activation frequency, increase within the first year of menopause.¹⁶ This bone loss is associated with significant trabecular perforations,¹⁷ which in themselves cause a dramatic reduction in bone strength and a predisposition to first fractures. The fact that circulating FSH levels are high during this period of rapid bone loss, even when estrogen levels have not declined, prompted us to investigate whether FSH directly affects the skeleton.

Stimulation of Bone Resorption by Follicle-Stimulating Hormone

We reported that recombinant FSH stimulated osteoclast formation from human or mouse bone marrow precursors in the presence of RANK-L.¹⁸ This finding has recently been confirmed by others.¹⁹ In addition, we showed that FSH stimulated the resorption of bone by mature human osteoclasts.¹⁸

The osteoclastogenic and proresorptive actions of FSH are exerted through a $G_{i2\alpha}$ -coupled

follicle-stimulating hormone receptor (FSHR), which we identified on both human and mouse osteoclasts and osteoclast precursors using reverse-transcription polymerase chain reaction, western blotting, immune staining, and flow cytometry. We also found that FSH enhanced the phosphorylation of downstream RANK-L-sensitive kinases, namely Erk and Akt, as well as $I\kappa B\alpha$, an inhibitor of nuclear factor kappa B; all of these pathways are known to transduce the proresorptive effects of RANK-L.¹⁸ Together, these studies have established a link between FSHR activation, $G_{i2\alpha}$ coupling, and kinase phosphorylation in mediating the effects of FSH.

Although we have documented direct effects of FSH on osteoclast signaling pathways,¹⁸ FSH also stimulates the production of TNF α from bone marrow precursor macrophages.²⁰ We have not studied FSH-induced increases in TNF α from T cells, and it is quite possible that T cells express FSHRs (Roberto Pacifici, Emory University, personal communication, 2009). If this is the case, FSH may be the primary trigger through which estrogen deficiency increases TNF α production from T cells to contribute to the bone loss in hypogonadism.

Estrogen Effects Are Distinct from Follicle-Stimulating Hormone Effects In Vivo

The strength of our in vitro observations suggests that elevated FSH levels could contribute to the

The strength of our in vitro observations suggests that elevated FSH levels could contribute to the enhanced osteoclastogenesis and resorption that characterize hypogonadal states.

enhanced osteoclastogenesis and resorption that characterize hypogonadal states.

However, to provide genetic evidence for an effect of FSH on the skeleton that is exerted independently of estrogen, we used mice lacking the β subunit of FSH or the FSHR.¹⁸ Genetic evidence from FSH β mutant mice was compelling, in that haploinsufficiency (50% reduction) of circulating FSH in the heterozygotes resulted in a high bone mass, despite these mice being eugonadal.¹⁸ Importantly, we found that the high bone mass arose from reduced bone resorption evident upon histomorphometry and ex vivo cultures of bone marrow cells.¹⁸ The enhanced bone mass and osteoclast defect in

eugonadal but FSH halpoinufficient mice separated the skeletal actions of FSH and estrogen.

Hypogonadal FSH $\beta^{-/-}$ mice also failed to lose bone despite severe hypogonadism, as did FSHR $^{-/-}$ mice in our hands.¹⁸ Nonetheless, both homozygous strains failed to show an increase in resorption parameters in vivo and ex vivo.¹⁸ It has been recently demonstrated that FSHR $^{-/-}$ mice do show an age-dependent loss of bone mass, which is further reduced upon androgenic withdrawal.²¹ However, this reduction in bone mass is largely due to inhibitory effects of estrogen and androgen depletion on bone formation, with a remarkably small ($\sim 4\%$ – 5%) increase in bone resorption.¹⁸

Significant elevations in bone resorption do not occur in the absence of a high FSH level.

Significant elevations in bone resorption do not occur in the absence of a high FSH level.

In particular, aromatase $^{-/-}$ mice, a model for chronic hypogonadism, display a doubling of resorption surfaces in the presence of high FSH, whereas estrogen receptor (ER) $\alpha^{-/-}\beta^{-/-}$ or gonadotropin releasing hormone (GnRH)^{mut} hpg mice, in which estrogen deficiency is as severe but is not accompanied by high FSH levels, do not show the expected increases in bone resorption.^{22,23} Some of these mice do lose bone, but this loss is due to reduced bone formation rather than elevated bone resorption.^{19,23} The exception to this is the deletion of estrogen receptor α in osteoclasts, in which bone loss and resorption increases do occur in the face of normal FSH levels; while establishing an effect of estrogen on the skeleton, which we do not dispute, these data do not rule out an additional action of FSH.²⁴

Thus, estrogen has a primary and profound anabolic action in these mouse models. That said, we are by no means excluding an antiresorptive effect of estrogen, the withdrawal of which in humans treated with gonadotropin-releasing hormone agonists causes high-turnover bone loss despite reduced FSH levels.²⁵

Our in vitro and in vivo evidence for distinct effects of estrogen and FSH on bone, in which FSH is proresorptive and estrogen is both anabolic and antiresorptive, creates a paradigm shift in our understanding of hypogonadal and, more importantly, perimenopausal and postmenopausal bone loss in women. We attribute bone loss in the early years of menopause not only to reduced estrogen but also to elevated FSH levels. This, we propose, assumes

particular relevance during late perimenopause when estrogen levels are normal but bone loss is still most profound.

Effect of Follicle-Stimulating Hormone on the Human Skeleton

Several clinical observations attest to a role of FSH in bone loss across the perimenopausal transition. The Study of Women Across Nations (SWAN), a cross-sectional and longitudinal study of over 2300 perimenopausal women between 42 and 52 years of age, showed that not only was there a strong correlation between basal FSH and bone remodeling markers,²⁶ but a change in FSH levels over 4 years predicted a change in bone mass.¹⁵ Likewise, Xu *et al.*²⁷ showed a significant increase in the incidence of osteoporosis in a group of 689 native Chinese women whose serum FSH was in the third and fourth quartiles, recapitulating data from SWAN.

It is indeed difficult to establish causality from SWAN or the Chinese study, as a high FSH level is a more sensitive marker for hypogonadism than reduced estrogen; hence, this association may simply be correlative. However, a step toward establishing causality is a smaller study showing that, with similar estrogen levels, amenorrheic women having a mean serum FSH level of 35 IU/L had considerably greater bone loss than women whose mean serum FSH level was 8 U/L.²⁸ Thus, the mouse and human studies make a compelling argument for further investigations on a role for FSH in causing hypogonadal bone loss, which traditionally has been attributable solely to low estrogen.

PATHOPHYSIOLOGY OF THYROTOXIC OSTEOPOROSIS

Hyperthyroidism and Osteoporosis in Humans

Von Recklinghausen²⁹ (1890) first reported an association between multiple fractures and hyperthyroidism. Since then, both anecdotal experience and patient-oriented research have established clear clinical associations between thyrotoxicosis and high-turnover osteoporosis. An elegant review of the literature between 1966 and 1997 by Greenspan and Greenspan² concluded that overt hyperthyroidism, including Graves' disease, is associated with a loss of skeletal integrity and a high risk of hip fracture. Likewise, the therapeutic suppression of TSH for thyroid cancer, goiter, or thyroid nodules is associated with osteoporosis, which is more pronounced

in menopause.^{3,4} Osteoporosis also accompanies Graves' disease, in which excessive resorption occurs mainly at cortical bone sites.⁵⁻⁷

In contrast, thyroid hormone replacement therapy to normalize serum TSH levels in hypothyroid patients seems to have a minimal effect, if any, on bone. This raises the question of whether the detrimental effect of thyroid hormone excess is mediated by thyroid hormones themselves, a suppressed TSH level, or even other factors, such as thyroid-stimulating hormone receptor (TSHR) antibodies. This question is accentuated by recent studies demonstrating that prior thyrotoxicosis is not associated with an increased fracture risk and that curing hyperthyroidism can at least partly reverse osteoporosis.

Pharmacological Evidence for Thyroid Hormone Action on Bone

In vitro data showing direct stimulatory effects of thyroid hormones on bone resorption support the overwhelming clinical association between hyperthyroidism, osteoporosis, and high fracture risk.^{30,31} This bone loss is due to accelerated resorption that is not compensated by a coupled increase in bone formation. Both T₄ and T₃ stimulate osteoclastic bone resorption in vitro and in organ cultures by initially activating osteoblasts, which then release RANK-L. RANK-L couples osteoblastic activation with enhanced osteoclastic bone resorption.^{32,33} There is considerable evidence that other cytokines, such as TNF α and interleukin 6 (IL-6), also mediate the pro-osteoclastic effects of hyperthyroidism, although the cellular origin of these cytokines remains uncertain. This idea of a cytokine-mediated effect is, however, supported by the high TNF α and IL-6 levels found in human hyperthyroidism.³⁴

Although high thyroid hormones indirectly stimulate bone resorption via the osteoblast, a direct effect on osteoblastic bone formation per se is evident during growth and development.³⁵ Hypothyroid mice and humans both have skeletal runting, a phenotype that is likely mediated through the decreased production of insulin-like growth factor 1. In contrast, thyrotoxicosis advances skeletal growth in utero and causes premature ossification postnatally in TR β PV mice that is reminiscent of juvenile human hyperthyroidism.³⁶ Thus, thyroid hormones constitute a requisite anabolic signal for optimal skeletal growth and modeling during development but are catabolic to the mature skeleton.

Despite in vitro, ex vivo, and clinical evidence supporting a direct action of thyroid hormone on the skeleton, certain observations remain unexplained.

First, patients with subclinical hyperthyroidism, in whom serum thyroid hormones are normal but TSH is suppressed, display high-turnover osteoporosis, as if the thyroid hormones were actually elevated.³⁷⁻⁴⁰ Second, mice deficient in both α_1 and β thyroid hormone receptors (TRs) display hypothyroid runting rather than defects in bone remodeling.^{41,42} In fact, their bone mass is conserved and, in some cases, slightly elevated. Finally, both bone density and fracture risk correlate strongly with serum TSH levels and not with serum thyroid hormones.^{43,44}

Genetic Evidence for Thyroid Hormone Action on Bone

A recent report dissects the role of the α TRs versus the β TRs in bone mass regulation.^{45,46} TR α null mice with normal thyroid hormone and TSH levels displayed high bone mass. This study demonstrates a role for TR α in bone mass regulation but currently without an examination of the precise cellular defect. Notably, it is unclear whether the high bone mass in TR α -deficient mice results from enhanced osteoblastic bone formation, reduced osteoclastic bone resorption, or a combination of both effects.

Even more remarkable was the finding that TR β deficient mice, in which both TSH and thyroid hormone were elevated because of the loss of thyroid hormone feedback on TSH release, displayed bone loss.⁴⁵ This unequivocal confirmation of thyroid hormone action on bone does not exclude a protective effect of elevated TSH. The effect of thyroid hormone could indeed be so profound that any osteoclast-inhibitory action of TSH might have failed to protect against bone loss. Alternately, a high circulating TSH level could inhibit bone formation (as discussed next).

Thyroid-Stimulating Hormone Directly Suppresses Bone Remodeling: Genetic Evidence

We reported that TSH is a direct inhibitor of both osteoclasts and osteoblasts.⁴⁷ An accompanying

We reported that TSH is a direct inhibitor of both osteoclasts and osteoblasts.

editorial classified it as "the bone suppressing hormone" because haploinsufficiency of TSHR in heterozygote mice triggered high-turnover osteoporosis in the face of normal thyroid function.⁴⁸ Furthermore, TSHR^{-/-} mice were also osteoporotic; this

phenotype in a hypothyroid state could not possibly arise from the proresorptive actions of thyroid hormones. Thyroid hormone replacement in TSHR^{-/-} mice failed to restore bone mass but corrected the runting. These data suggested that the effects of TSH on the skeleton were independent of circulating thyroid hormone concentrations. Furthermore, these studies brought into context earlier reports of TSHR expression in bone cells, which was also confirmed by Bassett and colleagues,⁴⁵ and attributed, at least by implication, the osteoporosis of hyperthyroidism to low TSH levels rather than solely to high thyroid hormones.⁴⁹

Data from Abe *et al.*,⁴⁷ Hase *et al.*,⁵⁰ and Sun *et al.*⁵¹ underpin evidence that TSH inhibits bone resorption directly. TSHR^{-/-} mice display evidence of enhanced osteoclast formation, as do hypothyroid (*byt/byt*) mice that have defective TSHR signaling.^{47,51} As expected from these data, recombinant TSH attenuates the genesis, function, and survival of osteoclasts *in vitro*.⁴⁷ Likewise, overexpression of the constitutively activated TSHR in osteoclast precursor RAW.C3 cells⁵⁰ or its transgenic expression in mouse precursors⁵¹ inhibits osteoclastogenesis. This inhibition can be explained by a reduction in nuclear factor kappa B and c-Jun N-terminal kinase signaling as well as attenuated production of the osteoclastogenic cytokine TNF α .^{47,50} TNF α production is up-regulated in TSHR^{-/-} mice, and this, we believe, increases osteoclast formation, contributing, at least in part, to the osteopenic phenotype. Finally, the genetic deletion of TNF α in TSHR^{-/-} mice therefore rescues the increased osteoclastogenesis in the double homozygotes.⁵⁰ The evidence from such genetic and pharmacological studies showing that TSH signaling is inhibitory to osteoclast formation is thus compelling.

Osteoblastic effects of TSH are, however, somewhat less clear, except that the increased osteoblastogenesis noted in TSHR deficiency is associated with the elevated expression of low-density lipoprotein receptor-like protein 5 and fetal liver kinase 1 (vascular endothelial growth factor receptor) proteins.⁴⁷ Recombinant TSH expectedly inhibits expression of these receptors, but there are surprisingly no alterations in the expression of 2 key osteoblastic transcription factors, osterix and runt-related transcription factor 2.⁴⁷

Exogenously Administered Thyroid-Stimulating Hormone Suppresses Bone Resorption

Evidence for the direct suppression of bone resorption by TSH is mounting,^{51,52} although it seems

Evidence for the direct suppression of bone resorption by TSH is mounting, although it seems that intermittently administered TSH may also be anabolic, despite a possible anti-anabolic action of the same hormone when it circulates constantly at high levels.

that intermittently administered TSH may also be anabolic,⁵⁰ despite a possible anti-anabolic action of the same hormone when it circulates constantly at high levels.

That is, in postmenopausal women, a single subcutaneous injection of recombinant human TSH sharply reduces serum crosslaps within 2 days to premenopausal levels, and this is followed by recovery after 7 days.⁵² In rats, TSH administered 3 times a week inhibits ovariectomy-induced bone loss and restores the lost bone 28 weeks after ovariectomy.⁵³ The latter observations on the restoration of lost bone have been attributed to anabolic effects of intermittent TSH due to increases in bone formation rates in calcein-labeled studies.⁵⁴ In a subsequent study,⁵¹ it was shown that, similarly to treatment with estrogen or intermittent parathyroid hormone, TSH could prevent ovariectomy-induced bone loss in mice. We also found that bone loss after 28 weeks of ovariectomy in rats could be restored by the injection of TSH once a week or once every 2 weeks; this was a remarkable finding in light of the short half-life of rat TSH.

Clinical studies on the effect of recombinant TSH in humans on serum bone turnover marker levels are just beginning to emerge. Recombinant human TSH as a single injection was shown to reduce serum crosslaps.⁵² However, effects on serum RANK-L and osteoprotegerin levels have been controversial. Although Giusti and coworkers⁵⁴ showed no alterations in these cytokines in response to recombinant TSH in patients receiving thyroxine for differentiated thyroid carcinoma, Martini and coworkers⁵⁵ found an increase in RANK-L after an initial elevation of procollagen type 1 N-terminal propeptide, a marker of bone formation. The latter study suggests that recombinant TSH administered as a bolus initially affects bone formation, a conclusion that seemed evident from rat studies.⁵³

Note that in none of the studies on rodents^{51,53} or indeed humans⁵² was there an elevation of serum thyroid hormone levels, and this proves beyond a reasonable doubt that the skeletal actions of TSH

(1) are independent of those of thyroid hormones, (2) are exerted at relatively low concentrations, and (3) in some instances can occur upon intermittent administration.

Thyroid-Stimulating Hormone and Human Skeletal Homeostasis

Several reports have indicated that serum TSH and bone loss correlate. A recent study has shown a strong negative correlation in postmenopausal women between the serum crosslinked telopeptide of type 1 collagen, a bone resorption marker, and serum TSH levels, whereas there was no such correlation with serum thyroid hormone levels.⁵⁶ l-Thyroxine administered to patients with nodular thyroid disease reduced bone mineral density to a significantly greater extent in patients in whom TSH levels were suppressed in comparison with those in whom TSH remained within the normal range.⁵⁷ Likewise, in patients with differentiated thyroid carcinoma, there is a strong inverse correlation between serum TSH bone turnover markers.⁵⁸ It has been shown that these correlations are independent of serum free T₄ and T₃ levels as well as other parameters of bone metabolism.⁵⁸ High circulating cathepsin K levels, a yet unvalidated marker for bone resorption, have also been reported in patients on suppressive doses of thyroxine for thyroid cancer.⁵⁹

The Tromsø study reported recently that subjects below the 2.5th percentile and above the 97.5th percentile of serum TSH levels had significantly lower and higher bone mineral densities, respectively, whereas no such association was noted expectedly between serum TSH and bone mineral density at normal TSH levels.⁶⁰ In an extended epidemiological study using data from the third US National Health and Nutrition Examination Survey, the odds ratios for correlations between serum TSH and bone mass were found to be in the range of 2 to 3.4. There was, in particular, a strong positive correlation between serum TSH within the reference range and bone mass in both black and white American women.⁴⁴ Finally, in limited observations of patients treated with thyroxine, serum TSH has been correlated to a number of indices of bone microarchitecture with heel ultrasound.⁶¹

Compelling data come from Bauer and colleagues,⁴³ who demonstrated that there is a dramatic 4.5-fold and 3.2-fold increase in vertebral and nonvertebral fracture rates, respectively, if the serum TSH level is lowered below 0.1 IU/L. This epidemiological study forces us to exert clinical caution in the use of thyroid hormone supplementation to lower serum TSH. However, although it is clear that

serum TSH correlates with bone mineral density, bone turnover markers, and fracture risk, it remains difficult to establish a cause-effect relationship.

Thyroid-Stimulating Hormone Receptor Polymorphisms and Bone Mass

Allelic associations between TSHRs, serum TSH, and bone mass, at this point, constitute perhaps the most powerful evidence for causality. Patients harboring the TSHR-D727E polymorphism had high bone mass; this association disappeared after correction for body weight.⁵⁸ The same allelic association has been reported in postmenopausal women from the United Kingdom and in the Rotterdam study.^{62,63} Yet another polymorphism in the Korean population, namely T+140974TC, has been shown to be associated with high bone mineral density at the lumbar spine and femoral neck, particularly in patient groups with an elevated TSH level, which would be consistent with an antiresorptive action of TSH.⁶²

PARADIGM SHIFT IN UNDERSTANDING OSTEOPOROSIS

We propose a paradigm shift in endocrine physiology. That is, classic pituitary hormones—FSH and

We propose a paradigm shift in endocrine physiology. That is, classic pituitary hormones—FSH and TSH—act by design on a nonendocrine tissue, bone.

TSH—act by design on a nonendocrine tissue, bone. Embedded within textbooks for decades has been the hard fact that these releasing hormones stimulate master hormone secretion from target endocrine organs. An ambitious extrapolation of our concept is that a plurality of pituitary hormones act in concert as part of a pituitary-bone axis to regulate skeletal integrity in health and disease. When dysregulated master hormone levels during hypogonadism and hyperthyroidism cause altered pituitary hormones through hypothalamic feedback, the latter contribute to the skeletal loss.

Thus, 2 new pathophysiological paradigms emerge from our studies. The osteoporosis of hyperthyroidism, which was thought to be caused solely by elevated thyroid hormone levels, can now be attributed at least in part to lowered TSH levels, particularly in patients in whom thyroid hormone levels are normal. Likewise, the

The osteoporosis of hyperthyroidism, which was thought to be caused solely by elevated thyroid hormone levels, can now be attributed at least in part to lowered TSH levels, particularly in patients in whom thyroid hormone levels are normal. Likewise, the osteoporosis of menopause, which was thought to arise solely from declining estrogen levels, can now be attributed to elevated FSH levels, particularly during perimenopause when estrogen levels are normal and bone loss is rapid.

osteoporosis of menopause, which was thought to arise solely from declining estrogen levels, can now be attributed to elevated FSH levels, particularly during perimenopause when estrogen levels are normal and bone loss is rapid. From an evolutionary standpoint, it also appears from our mouse genetic studies that the pituitary-bone axis is more ancient than the pituitary-ovary or pituitary-thyroid axes. This assumption is supported by the fact that haploinsufficiency of FSH β or TSHRs affects the skeleton without affecting the ovaries or thyroid gland, respectively. Furthermore, exogenous TSH inhibits bone resorption in mice, rats, and humans without affecting serum thyroid hormone levels. More work on how alterations in pituitary hormones synergize or antagonize other hormonal changes needs to be done. A recent longitudinal study⁶⁴ has highlighted the complexities in hormonal fluctuations during the menopausal transition, and it is very likely that target bone cells sum these changes to affect a particular outcome.

New questions thus emerge. Can new diagnostic tests be devised to detect changes in the pituitary-bone axis, reflecting larger system derangements and guiding personalized therapy? Likewise, can new biological therapies be devised that exploit the pituitary-bone axis to prevent bone loss in hypogonadism, hyperthyroidism, and other medical conditions? Further studies in mice and humans should allow these bedside-bench-bedside translations.

ACKNOWLEDGMENT

The reported studies were supported by grants from the National Institutes of Health (to Mone Zaidi, Terry Davies, and Li Sun), the Department of Veterans Affairs (to Terry Davies and Harry C. Blair), and the American Federation of Aging Research (to Jameel Iqbal).

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Stepnick LS. The frequency of bone disease. In: McGowan JA, Raisz LG, Noonan AS, Elderkin AL, eds. *Bone Health and Osteoporosis: A Report of the Surgeon General*. Washington, DC: Office of the US Surgeon General; 2004; 68–87.
- Greenspan SL, Greenspan FS. The effect of thyroid hormone on skeletal integrity. *Ann Intern Med* 1999; 130: 750–758.
- Ross DS. Hyperthyroidism, thyroid hormone therapy, and bone. *Thyroid* 1994; 4: 319–326.
- Baran DT. Thyroid hormone and bone mass: the clinician's dilemma. *Thyroid* 1994; 4: 143–144.
- Albright F, Smith PH, Richardson AM. Post menopausal osteoporosis. *JAMA* 1941; 1167: 2465–2474.
- Riggs BL, Khosla S, Melton LJ III. Sex steroids and the construction and conservation of the adult skeleton. *Endocr Rev* 2002; 23: 279–302.
- Shevde NK, Bendixen AC, Dienger KM, Pike JW. Estrogens suppress RANK ligand-induced osteoclast differentiation via a stromal cell independent mechanism involving c-jun repression. *Proc Natl Acad Sci U S A* 2000; 97: 7829–7834.
- Srivastava S, Toraldo G, Weitzmann MN, et al. Estrogen decreases osteoclast formation by downregulating receptor activator for NF-kappa B ligand induced JNK activation. *J Biol Chem* 2002; 276: 8836–8840.
- Srivastava S, Weitzmann MN, Kimble RB, et al. Estrogen blocks M-CSF gene expression and osteoclast formation by regulating phosphorylation of Egr-1 and its interaction with Sp-1. *J Clin Invest* 1998; 102: 1850–1859.
- Roggia C, Gao Y, Cenci S, et al. Up-regulation of TNF producing T cells in the bone marrow: a key mechanism by which estrogen deficiency induces bone loss in vivo. *Proc Natl Acad Sci U S A* 2001; 98: 13960–13965.
- García A Palacios V, Robinson LJ, Borysenko CW, et al. Negative regulation of RANKL induced osteoclastic differentiation in RAW264.7 cells by estrogen and phytoestrogens. *J Biol Chem* 2005; 280: 13720–13727.
- Armour KE, Armour KJ, Gallagher ME, et al. Defective bone formation and anabolic response to exogenous estrogen in mice with targeted disruption of endothelial nitric oxide synthase. *Endocrinology* 2001; 142: 760–766.

13. Jagger CJ, Chow JW, Chambers TJ. Estrogen suppresses activation but enhances formation phase of osteogenic response to mechanical stimulation in rat bone. *J Clin Invest* 1996; 98: 2351–2357.
14. Randolph JF, Sowers M, Bondarenko IV, et al. Change in estradiol and follicle-stimulating hormone across the early menopausal transition: effects of ethnicity and age. *J Clin Endocrinol Metab* 2004; 89: 1555–1561.
15. Sowers MR, Jannausch M, McConnell D, et al. Hormone predictors of bone mineral density changes during the menopausal transition. *J Clin Endocrinol Metab* 2006; 91: 1261–1267.
16. Recker R, Lappe J, Davies KM, Heaney R. Bone remodeling increases substantially in the years after menopause and remains increased in older osteoporosis patients. *J Bone Miner Res* 2004; 19: 1628–1633.
17. Akhter MP, Lappe JM, Davies KM, Recker RR. Transmenopausal changes in the trabecular bone structure. *Bone* 2007; 41: 111–116.
18. Sun L, Peng Y, Sharrow AC, et al. FSH directly regulates bone mass. *Cell* 2006; 125: 247–260.
19. Wu Y, Torchia J, Yao W, et al. Bone microenvironment specific roles of ITAM adapter signaling during bone remodeling induced by acute estrogen deficiency. *PLoS One* 2007; 2: e586.
20. Iqbal J, Sun L, Kumar TR, et al. Follicle stimulating hormone stimulates TNF production from immune cells to enhance osteoblast and osteoclast formation. *Proc Natl Acad Sci U S A* 2006; 103: 14925–14930.
21. Gao J, Tiwari-Pandey R, Samadfam R, et al. Altered ovarian function affects skeletal homeostasis independent of the action of follicle stimulating hormone. *Endocrinology* 2007; 148: 2613–2621.
22. Miyaura C, Toda K, Inada M, et al. Sex- and age-related response to aromatase deficiency in bone. *Biochem Biophys Res Commun* 2001; 280: 1062–1068.
23. Sims NA, Dupont S, Krust A, et al. Deletion of estrogen receptors reveals a regulatory role for estrogen receptor-beta in bone remodeling in females but not in males. *Bone* 2002; 30: 18–25.
24. Nakamura T, Imai Y, Matsumoto T, et al. Estrogen prevents bone loss via estrogen receptor alpha and induction of Fas ligand in osteoclasts. *Cell* 2007; 130: 811–823.
25. Dawson-Hughes B. Bone loss accompanying medical therapies. *N Engl J Med* 2001; 345: 989–992.
26. Sowers MR, Greendale GA, Bondarenko J, et al. Endogenous hormones and bone turnover markers in pre- and perimenopausal women: SWAN. *Osteoporos Int* 2003; 14: 191–197.
27. Xu ZR, Wang AH, Wu XP, et al. Relationship of age-related concentrations of serum FSH and LH with bone mineral density, prevalence of osteoporosis in native Chinese women. *Clin Chim Acta* 2009; 400: 8–13.
28. Devleta B, Adem B, Seneda S. Hypergonadotropic amenorrhea and bone density: new approach to an old problem. *J Bone Miner Res* 2004; 22: 360–364.
29. Von Recklinghausen F. *Die Fibrose Oder Deformierende Ostitis, die Osteomalazie und die Osteoplastische Carzinose in Ihren Gegenseitigen Beziehungen*. Berlin, Germany: George Reimer; 1891.
30. Murphy E, Williams GR. The thyroid and the skeleton. *Clin Endocrinol* 2004; 61: 285–298.
31. Wenzel KW. Thyroid hormone and bone. *Thyroid* 1994; 4: 508–509.
32. Allain TJ, Chambers TJ, Flanagan AM, McGregor AM. Tri-iodothyronine stimulates rat osteoclastic bone resorption by an indirect effect. *J Endocrinol* 1992; 133: 327–331.
33. Britto JM, Fenton AJ, Holloway WR, Nicholson GC. Osteoblasts mediate thyroid hormone stimulation of osteoclastic bone resorption. *Endocrinology* 1994; 134: 169–176.
34. Kim CH, Kim HK, Shong YK, et al. Thyroid hormone stimulates basal and interleukin (IL)-1-induced IL-6 production in human bone marrow stromal cells: a possible mediator of thyroid hormone-induced bone loss. *J Endocrinol* 1999; 160: 97–102.
35. Ishikawa Y, Genge BR, Wuthier RE, Wu LN. Thyroid hormone inhibits growth and stimulates terminal differentiation of epiphyseal growth plate chondrocytes. *J Bone Miner Res* 1998; 13: 1398–1411.
36. Bassett JH, Nordström K, Boyde A, et al. Thyroid status during skeletal development determines adult bone structure and mineralization. *Mol Endocrinol* 2007; 21: 1893–1904.
37. De Menis E, Da Rin G, Roiter I, et al. Bone turnover in overt and subclinical hyperthyroidism due to autonomous thyroid adenoma. *Horm Res* 1992; 37: 217–220.
38. Kisakol G, Kaya A, Gonen S, Tunc R. Bone and calcium metabolism in subclinical autoimmune hyperthyroidism and hypothyroidism. *Endocr J* 2003; 50: 657–661.
39. Földes J, Tarján G, Szathmari M, et al. Bone mineral density in patients with endogenous subclinical hyperthyroidism: is this thyroid status a risk factor for osteoporosis? *Clin Endocrinol* 1993; 39: 521–527.
40. Gurlek A, Gedik O. Effect of endogenous subclinical hyperthyroidism on bone metabolism and bone mineral density in premenopausal women. *Thyroid* 1999; 9: 539–543.
41. Kindblom JM, Gevers EF, Skrtic SM, et al. Increased adipogenesis in bone marrow but decreased bone mineral density in mice devoid of thyroid hormone receptors. *Bone* 2005; 36: 607–616.
42. Göthe S, Wang Z, Ng L, et al. Mice devoid of all known thyroid hormone receptors are viable but exhibit disorders of the pituitary-thyroid axis, growth, and bone maturation. *Genes Dev* 1999; 13: 1329–1341.
43. Bauer DC, Ettinger B, Nevitt MC, et al. Risk for fracture in women with low serum levels of thyroid-stimulating hormone. *Ann Intern Med* 2001; 134: 561–568.
44. Morris MS. The association between serum thyroid-stimulating hormone in its reference range and bone status in postmenopausal American women. *Bone* 2007; 40: 1128–1134.
45. Bassett JH, O'Shea PJ, Srisankantharajah S, et al. Thyroid hormone excess rather than thyrotropin deficiency induces osteoporosis in hyperthyroidism. *Mol Endocrinol* 2007; 21: 1095–1107.
46. Bassett D, Williams G. The skeletal phenotypes of TRalpha and TRbeta mutant mice. *J Mol Endocrinol* 2008; 42: 269–282.
47. Abe E, Mariani RC, Yu W, et al. TSH is a negative regulator of skeletal remodeling. *Cell* 2003; 115: 151–162.
48. Novack D. TSH, the bone suppressing hormone. *Cell* 2003; 115: 129–130.
49. Zaidi M, Sun L, Davies TF, Abe E. Low TSH triggers bone loss: fact or fiction? *Thyroid* 2006; 16: 1075–1076.

50. Hase H, Ando T, Eldeiry L, et al. TNFalpha mediates the skeletal effects of thyroid-stimulating hormone. *Proc Natl Acad Sci U S A* 2006; 103: 12849–12854.
51. Sun L, Vukicevic S, Baliram R, et al. Intermittent recombinant TSH injections prevent ovariectomy-induced bone loss. *Proc Natl Acad Sci U S A* 2008; 105: 4289–4294.
52. Mazziotti G, Sorvillo F, Piscopo M, et al. Recombinant human TSH modulates in vivo C-telopeptides of type-I collagen and bone alkaline phosphatase, but not osteoprotegerin production in postmenopausal women monitored for differentiated thyroid carcinoma. *J Bone Miner Res* 2005; 20: 480–486.
53. Sampath TK, Simic P, Sendak R, et al. Thyroid-stimulating hormone restores bone volume, microarchitecture, and strength in aged ovariectomized rats. *J Bone Miner Res* 2007; 22: 849–859.
54. Giusti M, Cecoli F, Ghiara C, et al. Recombinant human thyroid stimulating hormone does not acutely change serum osteoprotegerin and soluble receptor activator of nuclear factor-kappa beta ligand in patients under evaluation for differentiated thyroid carcinoma. *Hormones (Athens)* 2007; 6: 304–313.
55. Martini G, Gennari L, De Paola V, et al. The effects of recombinant TSH on bone turnover markers and serum osteoprotegerin and RANKL levels. *Thyroid* 2008; 18: 455–460.
56. Zofkova I, Hill M. Biochemical markers of bone remodeling correlate negatively with circulating TSH in postmenopausal women. *Endocr Regul* 2008; 42: 121–127.
57. La Vignera S, Vicari E, Tumino S, et al. L-thyroxin treatment and post-menopausal osteoporosis: relevance of the risk profile present in clinical history. *Minerva Ginecol* 2008; 60: 475–484.
58. Heemstra KA, van der Deure WM, Peeters RP, et al. Thyroid hormone independent associations between serum TSH levels and indicators of bone turnover in cured patients with differentiated thyroid carcinoma. *Eur J Endocrinol* 2008; 159: 69–76.
59. Mikosch P, Kersch-Schindl K, Woloszczuk W, et al. High cathepsin K levels in men with differentiated thyroid cancer on suppressive L-thyroxine therapy. *Thyroid* 2008; 18: 27–33.
60. Grimnes G, Emaus N, Joakimsen RM, et al. The relationship between serum TSH and bone mineral density in men and postmenopausal women: the Tromsø study. *Thyroid* 2008; 18: 1147–1155.
61. Nagata M, Suzuki A, Sekiguchi S, et al. Subclinical hypothyroidism is related to lower heel QUS in postmenopausal women. *Endocr J* 2007; 54: 625–630.
62. Van der Deure WM, Uitterlinden AG, Hofman A, et al. Effects of serum TSH and FT4 levels and the TSHR Asp727Glu polymorphism on bone: the Rotterdam Study. *Clin Endocrinol (Oxf)* 2008; 68: 175–181.
63. Albagha OME, Natarajan R, Reid DM, et al. The D272E polymorphism of the human thyroid stimulating hormone receptor is associated with bone mineral density and bone loss in women from the UK. *J Bone Miner Res* 2005; 20: S341.
64. Rannevik G, Jeppsson S, Johnell O, et al. A longitudinal study of the perimenopausal transition: altered profiles of steroid and pituitary hormones, SHBG and bone mineral density. *Maturitas* 2008; 61: 67–77.

SPECIAL FEATURE

Controversies in the Management of Alcoholic Liver Disease

Hui-Hui Tan, MBBS, MRCP,¹ Sharad Virmani, MD,² and Paul Martin, MD³

¹Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore

²Department of Medicine, University of Miami School of Medicine, Miami, FL

³Division of Hepatology, Center for Liver Diseases, University of Miami School of Medicine, Miami, FL

OUTLINE

ALCOHOLIC LIVER DISEASE

Alcoholic Hepatitis

Liver Histology

Prognostic Indicators

TREATMENT OF ALCOHOLIC LIVER DISEASE

Abstinence

Treatment of Alcoholic Hepatitis

LIVER TRANSPLANTATION FOR ALCOHOLIC LIVER DISEASE

CONCLUSION

ABSTRACT

Alcohol is a risk factor for chronic disease burden in developed countries. Alcoholic liver disease affects 1% of the North American population and is the second most frequent indication for liver transplantation in the United States. It is a spectrum that ranges from simple hepatic steatosis to alcoholic hepatitis to steatohepatitis and eventually cirrhosis. The clinical spectrum of alcoholic hepatitis is wide and ranges from the asymptomatic patient to overt liver failure and death. Liver biopsy as a means of prognostication in alcoholic hepatitis has mostly been replaced with less invasive scoring systems. The management of alcoholic liver disease is challenging. Abstinence is the cornerstone of therapy and should include

rehabilitation with a multidisciplinary approach. No specific treatment is required in mild to moderate alcoholic hepatitis. In patients with severe hepatitis, there appears to be a moderate survival benefit from the use of either corticosteroids or pentoxifylline in the absence of contraindications to their use. Non-responders should have steroid therapy withdrawn by day 7, as persistence with therapy is not beneficial. Orthotopic liver transplantation remains the definitive therapy for decompensated alcoholic cirrhosis despite alcohol abstinence. More studies are needed to define the optimal timing of orthotopic liver transplantation and patients at risk of alcohol relapse post-transplant. *Mt Sinai J Med* 76:484–498, 2009. © 2009 Mount Sinai School of Medicine

Key Words: alcohol abuse, alcohol tolerance, alcoholic hepatitis, alcoholic liver disease, liver transplant.

Alcohol is ranked as the third most frequent risk factor for chronic disease burden in developed countries, being responsible for a 9.2% reduction in healthy life years.¹ Almost two-thirds of adult Americans drink some alcohol, the majority of whom drink light or moderate amounts without apparent ill effect.² Alcohol abusers (or problem drinkers) suffer negative consequences from drinking (such as unemployment, loss of family, accidental injury, or death).³ Alcoholism is a physical dependence (a withdrawal syndrome appears upon the interruption of drinking that is relieved by alcohol itself) that includes tolerance (increasing amounts of alcohol are required to produce the desired effects), impaired control (alcohol intake cannot be regulated once drinking has begun on any drinking occasion), and craving (dysphoria of abstinence, which leads to relapse).⁴

Alcoholic liver disease (ALD) affects approximately 1% of the North American population. However, its true prevalence is not known as mild

Address Correspondence to:

Hui-Hui Tan

Department of
Gastroenterology and
Hepatology Singapore General
Hospital Singapore
Email: mhyhui@gmail.com

disease is asymptomatic and does not come to medical attention. The prevalence differs geographically, with alcohol abuse an important global health issue.⁵ ALD accounts for more than 12,000 deaths per year and 300,000 years of potential life lost in the United States.⁶ It is one of the 10 most frequent causes of death in developed countries, being responsible for 3% (1.8 million) of all deaths.⁷ Age-adjusted death rates from alcohol-induced liver disease indicate that it accounts for 40% of deaths from cirrhosis and 28% of all deaths from liver disease.⁸ ALD is the second most frequent indication for liver transplantation in the United States.⁶

ALCOHOLIC LIVER DISEASE

ALD is a spectrum that ranges from simple hepatic steatosis to alcoholic hepatitis to steatohepatitis and eventually cirrhosis. Simple hepatic steatosis is the commonest histological finding and occurs in 90% of heavy drinkers but is rapidly reversible with abstinence.⁹

Epidemiological data suggest that a threshold of 80 g of daily alcohol in a male and 20 g in a female for an average of 10 to 12 years is necessary for alcoholic injury to develop.^{10,11} However, it is not clear why only a minority of individuals who ingest alcohol in excess develop ALD or progress to cirrhosis.¹² Synergistic factors such as chronic hepatitis C, obesity, and genetic factors may accelerate the development of ALD even at lower doses of alcohol consumption.

Alcoholic hepatitis occurs in up to 35% of heavy drinkers⁹ and is usually a precursor of cirrhosis. It is a clinicopathological syndrome that denotes hepatocellular necrosis and inflammation in the chronic alcoholic patient. The clinical spectrum is wide and ranges from the asymptomatic patient to overt liver failure and the risk of death. It is usually accompanied by a low-grade fever, jaundice, leukocytosis, and a mild elevation in liver enzymes.

ALCOHOLIC HEPATITIS

Liver Histology

Liver biopsy as a means of prognostication in alcoholic hepatitis has mostly been replaced with less invasive scoring systems. The correlations between the degree of steatosis and clinical liver function and between the histological findings and indication for treatment of alcoholic hepatitis are unclear. By international consensus, obligatory features on histology include the presence of parenchymal necrosis, Mallory bodies, and a neutrophilic infiltrate

in a perivenular distribution.^{13,14} Nonobligatory features that are commonly present are bridging necrosis, fatty changes, bile duct proliferation, cholestasis, and perivenular fibrosis. It is reasonable to perform liver biopsy if the diagnosis of alcoholic hepatitis is in doubt or if a concomitant pathology (eg, hepatitis C) is suspected.¹⁵ Alcoholics with hepatitis C have a higher mortality risk than patients with ALD alone,^{16,17} as do patients with cirrhosis and alcoholic hepatitis in comparison with patients with only cirrhosis.^{18,19}

Liver biopsy as a means of prognostication in alcoholic hepatitis has mostly been replaced with less invasive scoring systems.

Prognostic Indicators

The Maddrey discriminant function (DF) score remains the most commonly used predictive model and was developed to facilitate the assessment of response in a clinical trial of steroids in patients with alcoholic hepatitis (Table 1).²⁰ A modified DF score >32 in the presence of hepatic encephalopathy predicts >50% mortality within 28 days in alcoholic hepatitis^{21,22}:

$$\text{Modified DF} = 4.6 (\text{Patient's PT} - \text{Control}) + \text{Serum bilirubin (mg/dL)}$$

where PT is the prothrombin time. However, fatal outcomes also occur in patients with modified DF scores <32, and its low specificity has suggested a need for alternative scoring systems.²³

Although the Model for End-Stage Liver Disease (MELD) score was initially developed to predict survival in patients with cirrhosis undergoing transjugular intrahepatic portosystemic shunting,²⁴ it has since been revised and validated to predict survival for patients with cirrhosis and to be the basis for organ donor allocation in liver transplantation.^{25,26} Laboratory values used to calculate the MELD score²⁷ are the international normalized ratio, creatinine, and bilirubin. MELD and modified DF scores (calculated within 24 hours of presentation) are equivalent in predicting 30- and 90-day mortality in alcoholic hepatitis,²⁸⁻³⁰ although MELD may be a better predictor of survival when it is calculated after the first week of hospitalization.²⁹ A MELD score ≥ 21 (within 24 hours of presentation) is a predictor of 90-day mortality (sensitivity = 0.75, specificity = 0.75).²⁸

Table 1. *Prognostic Indicators for Alcoholic Hepatitis.*

Prognosticator	Parameter	Predictive Result
Modified DF ²¹	1.6 (Patient's PT – Control) + Serum bilirubin (mg/dL)	Modified DF > 32 with hepatic encephalopathy predicts >50% mortality within 28 days.
MELD ²⁴	$9.57 \times \log_e [\text{Cr (mg/dL)}] + 3.78 \times \log_e [\text{Bilirubin (mg/dL)}] + 11.20 \times \log_e (\text{INR}) + 6.43^*$	MELD ≥ 21 predicts 20% mortality in 90 days. ²⁸
GAHS ³¹	Score 1 2 3 Age <50 ≥ 50 – WCC ($10^9/\text{L}$) <15 ≥ 15 – Urea (mmol/L) <5 ≥ 5 – PT ratio <1.5 1.5–2.0 >2.0 Bilirubin ($\mu\text{mol/L}$) <125 125–250 >250	GAHS ≥ 9 predicts >50% mortality in 28 to 84 days.
Lille model ³³	$3.19 - 0.101 \times \text{Age (years)} + 0.147 \times \text{Albumin on day 0 (g/L)}$ $+ 0.0165 \times \text{Evolution in bilirubin level } (\mu\text{mol/L}) - 0.206 \times \text{Renal insufficiency}$ $- 0.0065 \times \text{Bilirubin on day 0 } (\mu\text{mol/L}) - 0.0096 \times \text{PT (seconds)}^\dagger$	Lille score ≥ 0.45 predicts 75% mortality within 6 months in patients who have received corticosteroid therapy.

Abbreviations: Cr, creatinine; DF, discriminant function; GAHS, Glasgow Alcoholic Hepatitis Score; INR, international normalized ratio; MELD, Model for End-Stage Liver Disease; PT, prothrombin time; WCC, white cell count.

*See <http://www.mayoclinic.org/meld/mayomodel5.html>.

†See <http://www.lillemodel.com>.

The Glasgow Alcoholic Hepatitis Score (GAHS) was developed in an effort to overcome the low specificity of the Maddrey DF and lack of an optimal predictive cutoff point for the MELD score. It was derived from a retrospective analysis of patients with ALD, with logistic regression used to identify variables associated with mortality.³¹ This was then validated with the modified DF and MELD scoring systems. GAHS is a scoring system based on age, serum bilirubin, blood urea nitrogen, PT, and the peripheral white cell count. GAHS ≥ 9 is a predictor of mortality and is more accurate than DF in predicting both 28- and 84-day mortality but is equivalent to MELD in predicting 28-day mortality.

The Lille model³² incorporates age, renal insufficiency, albumin, PT, bilirubin, and the evolution of bilirubin on day 7 to predict 6-month mortality in patients with severe alcoholic hepatitis who have received corticosteroid therapy.³³

A recent single-center analysis of the various prognostic scores for acute alcoholic hepatitis (Lille, Glasgow, and Maddrey scores) and cirrhosis (MELD, MELD-Na, and Child-Turcotte-Pugh) in 108 patients treated with corticosteroids found that a Lille score ≥ 0.45 and GAHS ≥ 9 were the most accurate models for the prediction of mortality.³⁴ Although the specifics of the scoring systems vary, each of them helps clinicians to identify patients with alcoholic hepatitis with a high likelihood of short-term mortality who should be considered for more than supportive medical care alone.

TREATMENT OF ALCOHOLIC LIVER DISEASE

Abstinence

Abstinence is the cornerstone of therapy in the management of ALD.³⁵

Abstinence is the cornerstone of therapy in the management of ALD.

Ideally, this includes rehabilitation with a multidisciplinary approach. If abstinence is achieved, clinical^{36,37} and histological^{38–40} benefits can be impressive, even if the patient is already cirrhotic. Drugs (eg, disulfiram,^{41–43} naltrexone,⁴⁴ acamprosate,^{45,46} and baclofen^{47,48}) used to induce or maintain abstinence either have not had consistent beneficial effects or have been studied only in small numbers of patients. Disulfiram inhibits the activity of acetaldehyde dehydrogenase and leads to the accumulation of acetaldehyde, which results in flushing, tachycardia, hypotension, dyspnea, vomiting, vertigo, anxiety, and headache within 30 minutes of alcohol ingestion. Naltrexone has the potential to cause hepatocellular injury. Hence, both disulfiram and naltrexone have black box warnings issued about their use, especially in patients who continue to drink.

Treatment of Alcoholic Hepatitis

Apart from general supportive measures for the acutely ill patient with alcoholic hepatitis, several specific interventions have been studied.

Glucocorticoids

Glucocorticoids are the most intensely studied and yet most hotly debated treatment for acute alcoholic hepatitis.

Glucocorticoids are the most intensely studied and yet most hotly debated treatment for acute alcoholic hepatitis.

As the pathogenesis of alcoholic hepatitis has been linked to the activity of both the innate and adaptive immune systems, the rationale for glucocorticoid use is to block cytotoxic and inflammatory pathways in alcoholic hepatitis. Glucocorticoids have been shown to decrease proinflammatory cytokines and intercellular cell adhesion molecule 1 expression and inhibit polymorphonuclear neutrophil activation, and they have demonstrated short-term histological improvement in the treatment of alcoholic hepatitis.^{49,50}

Results from trials of glucocorticoids for ALD have been variable and depend on the nature of the trial and the group of patients recruited as the study population.

Results from trials of glucocorticoids for ALD have been variable and depend on the nature of the trial and the group of patients recruited as the study population.

Even among glucocorticoid trials with beneficial results, enrolled subjects were heterogeneous with variable definitions of randomization and blinding and without homogeneous inclusion or exclusion criteria, used different types of corticosteroids, used different scoring systems, and had different lengths of interventions and different co-interventions. Some studies had several different treatment arms,⁵¹ which made it difficult to draw meaningful conclusions during analysis. Among the various controversies surrounding corticosteroid use in alcoholic hepatitis is the risk of infection in an already immunocompromised host. Some trials have demonstrated higher mortality in the glucocorticoid group^{51–54} compared to the placebo group. Associated with this higher mortality, a greater incidence of fungal infections among patients receiving glucocorticoids has been reported by some authors.⁵³ An early meta-analysis on this subject, published in 1990, demonstrated a

protective effect of glucocorticoids in higher quality trials.⁵⁵ This was especially so in studies that excluded patients with gastrointestinal bleeding but included those with hepatic encephalopathy. In another meta-analysis conducted by Christensen and Gluud,⁵⁶ no overall treatment benefit was found after they attempted to control for confounders. A subsequent re-analysis of the same 3 randomized, controlled trials in Christensen and Gluud's meta-analysis, which pooled raw data from more than 200 patients with modified DF ≥ 32 , found a 28-day survival benefit of glucocorticoids (85%) versus placebo (65%).⁵⁷ In patients with modified DF ≥ 32 , treatment with glucocorticoids improved short term (28-day) survival, with mortality decreasing from 35% in controls to 15% with steroids. Conversely, patients with modified DF < 32 had a $> 90\%$ survival rate without steroids. In this analysis, the number of patients needed to treat to save 1 patient was 5. Another meta-analysis of 15 trials with 721 randomized patients reported that the evidence in favor of glucocorticoids was based on heterogeneous trials of low quality.⁵⁸ However, this is in contradiction to a further meta-analysis of raw data from the last 5 randomized controlled trials (with a total of 387 patients) using corticosteroids in severe alcoholic hepatitis,⁵⁹ which found significantly higher 28-day survival in patients treated with corticosteroids. This benefit was mainly observed in patients classified as responders according to the Lille model (ie, patients who demonstrated an early change in serum bilirubin levels by day 7 of corticosteroid therapy).

The longer term benefit of steroids is also difficult to assess as the various trials had differing follow-up periods, and unless the patient abstains from alcohol completely, alcoholic hepatitis is likely to recur.⁶⁰ The survival benefit of corticosteroid therapy has not been found to persist beyond 1 year.⁶¹

Patient selection for steroid use needs to be better defined. For now, the data appear to support its use in patients with a modified DF ≥ 32 in the absence of an active infection or gastrointestinal bleeding. Conversely, as per the exclusion criteria of most trials and meta-analyses, patients with mild alcoholic hepatitis (modified DF < 32) and evidence of infection and gastrointestinal bleeding would not benefit from glucocorticoid treatment. Patients who do respond to steroid therapy should be treated with prednisolone (40 mg daily for 4 weeks and then tapered).⁴ A study by Mathurin *et al.*⁶² suggests that patients with an early fall in bilirubin levels (defined as serum bilirubin on day 7 being lower than that on day 1 of treatment) are significantly more likely to survive with glucocorticoid therapy. Hence, some authors now recommend that patients with alcoholic

hepatitis who do not demonstrate an early reduction in bilirubin levels (ie, on day 7 of treatment) should have corticosteroids withdrawn. This initial study by Mathurin and colleagues eventually led to the construction of the Lille model described earlier. A score ≥ 0.45 on the Lille model is highly predictive of 6-month mortality with 75% sensitivity.

Pentoxifylline.

Elevated tumor necrosis factor alpha (TNF α) levels have been found to be predictive of poor survival in alcoholic hepatitis.^{63,64} Pentoxifylline is a non-selective phosphodiesterase inhibitor and a TNF α suppressor.⁶⁵

In 1991, a study of pentoxifylline for severe alcoholic hepatitis (DF ≥ 32) reported a reduction in the development of hepatorenal syndrome and mortality in comparison with placebo.⁶⁶ A subsequent study of 101 patients from the same center supported the earlier findings, demonstrating a 40% reduction in mortality in comparison with placebo.⁶⁷ The number needed to treat to prevent 1 death was 4.7. However, in this study, there was no demonstrable improvement in routine liver function tests, liver histology, or survival apart from the development of hepatorenal syndrome.⁶⁸ In another study, 29 patients who did not respond to corticosteroids (identified by an absence of an early decline in bilirubin) were switched to pentoxifylline for 28 days and compared

Although some data suggest a benefit with pentoxifylline in alcoholic hepatitis, it is unclear whether its benefit extends beyond possibly preventing hepatorenal syndrome.

to 58 other matched patients who persisted with corticosteroid therapy.⁶⁹ No survival benefit was observed with pentoxifylline at 2 months. Thus, although some data suggest a benefit with pentoxifylline in alcoholic hepatitis, it is unclear whether its benefit extends beyond possibly preventing hepatorenal syndrome.

Nutrition.

Nutrition trials in alcoholic hepatitis are heterogeneous with discernible benefits and remain controversial. Studies have varied in the types of subjects recruited, the nutritional intervention employed (eg, parenteral nutrition versus enteral nutrition and caloric amount and type), and the duration of intervention and have had inadequate control groups or

unbalanced control formulas. Study methods have even varied from replacing what was deficient in the alcoholic patient to providing less than the basic requirements of most hospitalized patients^{70,71} to true supplementation over and above basic nutritional requirements.⁷² Although some trials have reported some benefit with nutrition intervention (eg, improved nitrogen balance^{73,74} and improved liver function⁷⁵), most trials have failed to demonstrate a true survival benefit.

Although some trials have reported some benefit with nutrition intervention (eg, improved nitrogen balance and improved liver function), most trials have failed to demonstrate a true survival benefit.

Short-term clinical trials on the effects of oral nutritional supplementation showed tolerance to feeding without worsening of encephalopathy^{76–78}; one study showed improved liver function in Childs-Pugh class C patients.⁷⁸ In patients who voluntarily consumed more than 2500 kcal/day, oxandrolone combined with oral nutritional supplementation demonstrated an improvement in 6-month survival, liver function, and nutritional status in comparison with a placebo group.⁷⁹

Enteral tube feeding over a 4-week period was associated with significantly improved hepatic encephalopathy and aminopyrine clearance testing of liver function in comparison with alcoholic hepatitis patients on a conventional hospital diet.⁸⁰ A subsequent randomized, multicenter Spanish study demonstrated improved 1-year survival rates in alcoholic hepatitis patients who were tube-fed with an enteral nutrition formula containing 2000 kcal of enriched branched chain amino acids for 4 weeks in comparison with patients receiving prednisolone alone.⁸¹ In this study, there was no difference in overall mortality between the 2 groups during the 28-day treatment period.

An analysis of 7 studies suggested that parenteral nutrition treatment of up to 30 days may improve liver function and nitrogen balance; however, its long-term effects and benefits are uncertain.⁸²

Patients with ALD are malnourished⁸³ for a number of reasons, including malabsorption, the induction of a catabolic state, and the replacement of calories with alcohol.^{84–86} Protein-caloric malnutrition has also been demonstrated to correlate with short-term and long-term mortality in alcoholic

patients.⁸⁷ Hence, malnutrition should be actively sought in such patients, and replacement should be commenced accordingly. The American College of Gastroenterology guidelines recommend 1.2 to 1.5 g/kg of protein and 35 to 40 kcal/kg of body weight for patients with ALD.⁴

Anti-Tumor Necrosis Factor Alpha.

Infliximab (Remicade®) is an anti-TNF α mouse/human chimeric antibody and has been extensively studied in alcoholic hepatitis. Early reports were encouraging, demonstrating improved survival rates, improved Maddrey scores, or improved laboratory parameters.^{88–90} However, the largest randomized, controlled trial to date, which enrolled 36 patients and compared a combination of prednisolone (40 mg/day) and infliximab (10 mg/kg 3 times per week in weeks 0, 2, and 4) to prednisolone and placebo in alcoholic hepatitis was terminated prematurely. More deaths had occurred in the group treated with prednisolone and infliximab, although this finding did not reach statistical significance.⁹¹ Whether this risk was related to the higher doses of infliximab used, more ill patients being recruited in this trial in comparison with earlier studies, the combined use of prednisolone and infliximab, or the possible unsuitability of infliximab for the treatment of alcoholic hepatitis is still debatable. There have also been reports of acute liver failure in patients with Crohn's disease or rheumatoid arthritis treated with infliximab.⁹²

A small, uncontrolled study evaluated the use of etanercept (Enbrel®) in 13 patients with moderate to severe alcoholic hepatitis. Two patients died within 32 days, and 3 developed serious adverse events (infection, gastrointestinal bleeding, and hepatorenal syndrome) mandating withdrawal of the drug.⁹³ A more recent multicenter, randomized, double-blind, placebo-controlled study of etanercept in 48 patients with severe alcoholic hepatitis (defined as MELD ≥ 15) found no difference in the 1-month mortality rates in the 2 groups on an intention-to-treat analysis.⁹⁴ Alarming, the 6-month mortality in the etanercept group was significantly higher in comparison with the controls (57.7% versus 22.7%; odds ratio, 4.6; 95% confidence interval, 1.3–16.4; $P = 0.017$). Rates of serious adverse infectious events were also significantly higher in the etanercept group in comparison with the placebo group, and this led the authors to conclude that etanercept is not effective for the treatment of moderate to severe alcoholic hepatitis and that approaches to neutralize serum TNF α in such patients are not warranted.

Antioxidant Cocktails.

Oxidative stress is a key factor in the pathogenesis of alcoholic hepatitis and is the rationale for the use of antioxidants. However, results have failed to show any convincing benefit from their use.

Phillips *et al.*,⁹⁵ who compared corticosteroids to an antioxidant cocktail (β -carotene, vitamins C and E, selenium, methionine, allopurinol, desferrioxamine, and *N*-acetylcysteine), reported inferior survival rates in comparison with corticosteroids at 30 days. However, this survival advantage was lost at 1-year follow-up. Another study stratified patients with severe alcoholic hepatitis by gender and corticosteroid treatment.⁹⁶ The active group received a loading dose of *N*-acetylcysteine (150 mg/kg followed by 100 mg/kg/day for 1 week) and daily doses of vitamins A and E, biotin, selenium, zinc, manganese, copper, magnesium, folic acid, and coenzyme Q for 6 months. Antioxidant therapy showed no benefit, either alone or in combination with corticosteroids. A more recent study of 87 patients with severe alcoholic hepatitis (defined as modified Maddrey DF ≥ 32) randomized patients to receive either corticosteroids with *N*-acetylcysteine infusion for 5 days or corticosteroids alone.⁹⁷ Although patients in the *N*-acetylcysteine group had better 1-month survivals, this effect did not persist at 3 and 6 months.

Colchicine.

Colchicine was initially employed for ALD because of its antifibrotic properties in rats.⁹⁸ However, a Cochrane review of 15 human trials with 1714 randomized subjects failed to demonstrate a treatment benefit in patients with alcoholic or nonalcoholic fibrosis or cirrhosis.⁹⁹ A large trial of 549 patients with alcoholic cirrhosis also failed to demonstrate any treatment benefit with long-term use (up to 6 years) of colchicine.¹⁰⁰

S-Adenosyl Methionine.

S-Adenosyl methionine (SAME) is a precursor of glutathione that theoretically might be protective in alcohol-induced liver injury.¹⁰¹ In a Cochrane database review of 9 randomized trials that combined a heterogeneous sample of 434 patients with ALD, SAME failed to show a survival benefit.¹⁰² Only 1 trial of 62 patients deemed to have adequate methodology and outcome reporting and good quality¹⁰³ suggested benefit (improved survival and delay to liver transplantation) with 2 years of SAME treatment for Child's class A and B alcoholic cirrhosis.

Propylthiouracil.

Early animal studies with propylthiouracil (PTU) suggested benefits in oxidative and ischemic liver injury.^{104,105} A meta-analysis of 6 randomized trials of 710 patients failed to demonstrate any treatment benefit (in terms of survival or liver histology) in patients with ALD. Moreover, PTU was also found to be associated with adverse events, such as leukopenia.¹⁰⁶ There have also been case reports of hepatitis and fulminant hepatic failure with PTU use.^{107–110}

Anabolic Steroids.

The first human trial for anabolic steroids was conducted as early as in 1947.¹¹¹ The theoretical basis for its use dates back to 1938, when an animal study demonstrated enlargement of cirrhotic rat livers following the administration of androgens.¹¹² In 2006, an updated Cochrane systemic review of 5 randomized trials of 499 patients concluded that there was no significant benefit of anabolic-androgenic steroids in mortality, liver-related mortality, complications of liver disease, or liver histology for patients with ALD.¹¹³

LIVER TRANSPLANTATION FOR ALCOHOLIC LIVER DISEASE

Orthotopic liver transplantation (OLT) remains the definitive therapy for decompensated cirrhosis due to ALD despite continued alcohol abstinence.

Orthotopic liver transplantation (OLT) remains the definitive therapy for decompensated cirrhosis due to ALD despite continued alcohol abstinence.

ALD is the second most common indication for liver transplantation in North America.¹¹⁴ Short-term (1- to 7-year) graft survival and patient survival remain at par with, if not superior to, survival with non-ALD if the patient remains abstinent.¹¹⁵ Bellamy *et al.*¹¹⁶ reported that in 123 patients who were transplanted, patient survival at 1, 5, and 7 years was 84%, 72%, and 64%, respectively. Graft survival was 81%, 66%, and 50%, respectively, over the same period. The 1- and 5-year patient and graft survival rates for all patients with cirrhosis were 86.9%/73.4% and 82.4%/67.4%, respectively.¹¹⁷ A recent study also confirmed impressive survival rates in ALD with 1-, 3-, and 5-year patient survival rates of 96.2%, 89.6%, and

84.4%, respectively, post-OLT.¹¹⁸ However, OLT for ALD patients continues to fuel controversy, including issues of recidivism, potentially poor compliance with postoperative care, and inherent biases against alcoholics, such as concern about using scarce organs for what is often perceived to be a self-inflicted disease.

Numerous cohorts of apparently well-selected recipients with ALD continue to have high alcohol recidivism rates.^{119–131}

Numerous cohorts of apparently well-selected recipients with ALD continue to have high alcohol recidivism rates.

DiMartini *et al.*'s¹³² prospective study of alcoholic recipients found that 22% had used some alcohol by the end of the first year post-OLT and 42% had by 5 years, of whom 26% had a binge drinking pattern. Such a wide range of relapse rates may stem from varying definitions of recidivism and methods of eliciting alcohol consumption data. Most studies addressing recidivism in the past 20 years have used the "any use" definition of alcohol relapse.¹³³ However, a return to drinking does not necessarily mean excessive drinking.¹²³ Furthermore, Fabrega *et al.*'s¹²³ report of patients who had returned to drinking revealed no decreased compliance with other medical care, including immunosuppressant therapy. Steatosis was the only histological finding in 3 of 8 relapsers. Yet, stratifying drinking patterns as "slips," episodic drinking separated by periods of abstinence, and harmful drinking has promoted more qualitative and quantitative endpoints in outcome. Phitzmann *et al.*¹³⁰ stratified relapsers into slips and harmful drinking, revealing significantly worse 5- and 10-year survival rates (69.5% and 20%) among "harmful" drinkers versus abstainers (90.3% and 81.5%).¹³⁰ Although other studies have suggested increased similar morbidity in relapsers,^{134,135} the impact on rejection and survival is conflicting in different patient groups.^{125,136–138}

A major focus in determining candidacy for liver transplantation in ALD has been identifying factors to predict posttransplant recidivism. Pre-transplant duration of abstinence from alcohol was the first predictive factor analyzed (discussed later). Other medical and social variables can be associated with high relapse rates, including tobacco consumption, noncompliance to follow-up clinic visits, and mental illness.^{139,140} DiMartini *et al.*¹⁴¹ showed prior nonalcoholic substance abuse, alcoholism in a first-degree relative, and a history of alcohol rehabilitation

Table 2. Predictors of Relapse and Literature Investigating Such Predictors.

Predictor	Type of Study	Predicts Relapse Post-OLT	n	Heavy Drinking	Relapse (%)	Follow-Up Median or Mean (Range) in Months	Reference
Lack of social support	R	Yes	300	8*	19	89 (24–180)	Pfzmann <i>et al.</i> (2007) ¹³⁰
	R	Yes	90	20	31	67(not stated)	Kelley <i>et al.</i> (2006) ¹³⁹
	R	No†	111	14	15	44 (40–48)	Jauhar <i>et al.</i> (2004) ¹²⁹
	P	Yes	61	Not stated	20	83 (30–112)	Gish <i>et al.</i> (2001) ¹⁴⁰
	R	No	46	7	53	25 (5–47)	Mackie <i>et al.</i> (2001) ¹³¹
	R	No	53	10	32	42 (1–100)	Pageaux <i>et al.</i> (1999) ¹⁵³
	P	Yes	91	Not stated	20	Not stated (12–72)	Coffman <i>et al.</i> (1997) ¹⁵⁴
	R	Yes	63	Not stated	21	49 (28–70)	Foster <i>et al.</i> (1997) ¹²⁶
Family history	R	Yes	45	Not stated	11	15 (12–24)	Lucey <i>et al.</i> (1992) ¹⁵⁵
	P	No	167	26	42	60 (not stated)	DiMartini <i>et al.</i> (2006) ¹³²
	R	No‡	90	20	31	67 (not stated)	Kelly <i>et al.</i> (2006) ¹³⁹
	R	Yes	111	14	15	44 (40–48)	Jauhar <i>et al.</i> (2004) ¹²⁹
	P	Yes	36	Not stated	38	Not stated (3–12)	DiMartini <i>et al.</i> (2001) ¹⁴¹
	R	No	46	24	54	25 (5–47)	Mackie <i>et al.</i> (2001) ¹³¹
	R	Yes	63	Not stated	21	49 (28–70)	Foster <i>et al.</i> (1997) ¹²⁶
	R	Yes	300	8*	19	89 (24–180)	Pfzmann <i>et al.</i> (2007) ¹³⁰
Other psychiatric diagnoses	R	Yes	90	20	31	67 (not stated)	Kelly <i>et al.</i> (2006) ¹³⁹
	R	No	111	14	15	44 (40–48)	Jauhar <i>et al.</i> (2004) ¹²⁹
	P	No	36	Not stated	38	Not stated (3–12)	DiMartini <i>et al.</i> (2001) ¹⁴¹
	P	Yes	61	Not stated	20	83 (30–112)	Gish <i>et al.</i> (2001) ¹⁴⁰
	P	Yes	91	Not stated	20	Not stated (12–72)	Coffman <i>et al.</i> (1997) ¹⁵⁴
	R	Yes	68	Not stated	13	Not stated (0–84)	Tripp <i>et al.</i> (1996) ¹⁵⁶
	P	Yes	61	Not stated	20	83 (30–112)	Gish <i>et al.</i> (2001) ¹⁴⁰
	P	Yes	36	Not stated	38	Not stated (3–12)	DiMartini <i>et al.</i> (2001) ¹⁴¹
Prior nonadherence to treatment	Case-control	No	56	16	50	24 (0–84)	Tang <i>et al.</i> (1998) ¹³⁶
	P	Yes	167	26	42	60 (not stated)	DiMartini <i>et al.</i> (2006) ¹³²
	R	No	90	20	31	67 (not stated)	Kelly <i>et al.</i> (2006) ¹³⁹
	R	No	111	14	15	44 (40–48)	Jauhar <i>et al.</i> (2004) ¹²⁹
	P	Yes	34	Not stated	38	Not stated (3–12)	DiMartini <i>et al.</i> (2001) ¹⁴¹
	P	No	91	Not stated	20	Not stated (12–72)	Coffman <i>et al.</i> (1997) ¹⁵⁴
	R	Yes	63	Not stated	21	49 (28–70)	Foster <i>et al.</i> (1997) ¹²⁶
	R	Yes	63	Not stated	21	49 (28–70)	Foster <i>et al.</i> (1997) ¹²⁶

NOTE: All studies had at least 30 patients.

Abbreviations: OLT, orthotopic liver transplantation; P, prospective cohort study; R, retrospective cohort study.

*5.5% had an “unknown” amount of drinking.

†Marginal predictor ($P < 0.07$) on univariate analysis; not stated on multivariate analysis.

‡Marginal predictor ($P = 0.051$).

to be independently associated with posttransplant relapse.¹⁴¹ Yet, there is marked interstudy variability when recidivism is being predicted (Table 2).¹⁴² These differences could stem from the definition of recidivism or the design of the studies involved. A 2008 meta-analysis of 50 studies looking at predictors of recidivism found 3 significant, albeit modest, variables: a poor social support system, a family history of alcohol abuse/dependence, and pretransplant abstinence of 6 months or less.¹²⁴ The authors further stated that most studies attempted to identify

factors of predictability in an independent fashion rather than cumulative fashion. Regardless, a well-rounded approach to each patient is necessary in assessing the medical, psychiatric, and social conditions for transplant candidacy. The question remains of how much weight to attach to each variable when posttransplantation recidivism is being predicted.

Pre-OLT abstinence, especially the 6-month rule, remains contentious. Studies over the years have provided convincing data for and against

Table 3. Literature on 6 Months of Abstinence Predicting Relapse.

Predictor	Type of Study	Predicts Relapse Post-OLT	n	Heavy Drinking (%)	Relapse (%)	Follow-Up Median or Mean (Range) in Months	Reference
Six months of alcohol abstinence	P	Yes	51	22	30	36 (1–86)	Miguet <i>et al.</i> (2004) ¹²⁵
	P	Yes	37	8	19	21 (not stated)	Osorio <i>et al.</i> (1994) ¹²²
	P	Not assessed†	167	26	42	60 (not stated)	DiMartini <i>et al.</i> (2006) ¹³²
	P	No	36	Not stated	38	Not stated (3–12)	DiMartini <i>et al.</i> (2001) ¹⁴¹
	P	No	61	Not stated	20	83 (30–112)	Gish <i>et al.</i> (2001) ¹⁴⁰
	P	No	34	12	33	40 (0–86)	Burra <i>et al.</i> (2000) ¹⁵⁷
	P	No	91	Not stated	20	Not stated (12–72)	Coffman <i>et al.</i> (1997) ¹⁵⁴
	R	Yes	300	8*	19	89 (24–180)	Pfützmann <i>et al.</i> (2007) ¹⁵⁰
	R	No	90	20	31	67 (not stated)	Kelly <i>et al.</i> (2006) ¹⁵⁹
	R	No	103	18	33	31 (6–60)	Bjornsson <i>et al.</i> (2005) ¹⁵⁸
	R	No	111	14	15	44 (40–48)	Jauhar <i>et al.</i> (2004) ¹²⁹
	R	No	123	Not stated	11	84 (50–96)	Bellamy <i>et al.</i> (2001) ¹¹⁶
	Meta-analysis	Yes	313	Not stated	32	Not applicable	Bravata <i>et al.</i> (2001) ¹⁵⁹
	R	No	46	7	53	25 (5–47)	Mackie <i>et al.</i> (2001) ¹³¹
	R	No	185	Not stated	20	94 (83–105)	Jain <i>et al.</i> (2000) ¹⁶⁰
	R	No†	56	27	50	30 (6–120)	Pereira <i>et al.</i> (2000) ¹⁶¹
	R	Yes	167	Not stated	26	24 (not stated)	Platz <i>et al.</i> (2000) ¹⁶²
	R	Yes	53	10	32	42 (1–100)	Pageaux <i>et al.</i> (1999) ¹⁴⁹
	R	No	78	Not stated	24	Not stated	DiMartini <i>et al.</i> (1998) ¹⁶³
	Case-control	No	56	16	50	24 (0–84)	Tang <i>et al.</i> (1998) ¹³⁶
	R	No‡	63	Not stated	21	49 (28–70)	Foster <i>et al.</i> (1997) ¹²⁶
	R	No	41	5	49	47 (36–95)	Gerhardt <i>et al.</i> (1996) ¹⁶⁴
	R	No	58	17	21	27 (18–46)	Tringali <i>et al.</i> (1996) ¹⁶⁵
	R	Yes§	57	Not stated	33	Not stated (not stated)	Doffoel <i>et al.</i> (1992) ¹⁶⁶
	R	Yes	52	Not stated	12	25 (16–34)	Kumar <i>et al.</i> (1990) ¹⁶⁷

NOTE: All studies had at least 30 patients.

Abbreviations: OLT, orthotopic liver transplantation; P, prospective cohort study; R, retrospective cohort study.

*5.5% had an “unknown” amount of drinking.

†The length of abstinence was significantly different between relapsers and nonrelapsers.

‡The length of abstinence was significantly different between relapsers and nonrelapsers; 6-month abstinence was not statistically predictive.

§Three-month abstinence was statistically predictive.

the 6-month abstinence requirement (Table 3). Most transplant centers, though not all, require patients to have a period of abstinence, usually for at least 6 months, and have proof of adequate social or caregiver support.¹⁴³ Lucey *et al.*¹⁴⁴ suggested that this 6-month period of abstinence would allow the native liver to recover with medical management and possibly obviate transplantation. However, this minimal period of abstinence is sometimes waived if the patient is deemed too ill to survive beyond 6 months without a liver transplant. One study showed that recovery in decompensated alcoholic cirrhosis by alcohol abstinence can be predicted

within 3 months of abstinence by the monitoring of clinical signs via the Child-Pugh scoring system (serum bilirubin, albumin, international normalized ratio, ascites, and hepatic encephalopathy).¹⁴⁵ Veldt *et al.*¹⁴⁵ found that although such improvement in liver function can take place within 3 months of abstinence, some abstinent patients die within 6 months; this has led some authors to suggest reducing the period of abstinence from 6 to 3 months.¹⁴⁶ Yet, abstinence less than 12 months was recently identified as a significant risk factor for relapse in a large retrospective study of liver transplant recipients.¹¹⁸ In practice, most centers do

not adhere strictly to a fixed period of abstinence but rather consider each case on an individual basis, balancing the psychosocial and clinical factors.¹⁴⁷ It is noteworthy that Kelly *et al.*'s study¹³⁹ did not find the period of abstinence to be predictive of post-liver transplant alcohol relapse.¹³⁹ Although excessive drinking has been reported to be associated with progressive liver injury and fatal alcoholic hepatitis,^{135,148} Wells *et al.*¹⁴⁹ found no difference in patient or graft survival when they compared patients with histological alcoholic hepatitis and patients with bland alcoholic cirrhosis who had undergone liver transplantation. Yet, most centers are reluctant to transplant patients with severe acute alcoholic hepatitis as this implies recent drinking and a risk of recidivism post-OLT. Most centers in the United States and insurance companies also require the patient to enter a formal alcoholic rehabilitation program and require patients to sign a contract to abstain from drinking.¹²⁰ A recent study highlighted the difficulty of advising OLT for at least some patients with ALD. It suggested that patients with Child-Pugh B cirrhosis did not necessarily have a survival benefit with immediate listing for OLT in comparison with persisting with medical therapy.¹⁵⁰

Finally, an increasing concern of late is the high risk for de novo malignancies in long-term survivors transplanted for ALD. Although posttransplantation lymphoproliferative disorder and non-melanoma skin cancer remain the most common malignancies post-liver transplantation, the incidence of esophageal cancer is significantly increased in patients with alcoholic etiologies. Duvoux *et al.*'s prospective study¹⁵¹ showed a significantly higher incidence of malignancies in patients with ALD compared to those with non-ALD etiologies. They also detected squamous cell carcinoma of the oropharynx or esophagus only in recipients transplanted for ALD. Risk factors undoubtedly include the cumulative effects of alcohol and, in most cases, smoking with posttransplant immunosuppression.¹⁵² Thus, regular ear, nose, and throat examinations appear justified in patients transplanted for ALD.

CONCLUSION

The management of ALD remains a challenge. Holistic management requires a multidisciplinary approach to facilitate abstinence and adequate nutrition. No specific treatment is required in mild to moderate alcoholic hepatitis. In patients with severe hepatitis, there appears to be a moderate survival benefit from the use of either corticosteroids

or pentoxifylline. Nonresponders should have steroid therapy withdrawn by day 7, as persistence with therapy has not been shown to be beneficial. Motivated patients who do not improve despite alcohol abstinence should be considered for liver transplantation. Further studies are needed to define the optimal timing of OLT and patients at risk of alcohol relapse post-transplant.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Tsukamoto H. Conceptual importance of identifying alcoholic liver disease as a lifestyle disease. *J Gastroenterol* 2007; 42: 603–609.
2. Midanik L, Room R. The epidemiology of alcohol consumption. *Alcohol Health Res World* 1992; 16: 183–190.
3. Gordis E. The etiology, consequences, and treatment of alcoholism. *Liver Transpl Surg* 1997; 3: 199–205.
4. McCullough AJ, O'Connor JF. Alcoholic liver disease: proposed recommendations for the American College of Gastroenterology. *Am J Gastroenterol* 1998; 93: 2022–2036.
5. The World Health Report 2002. Available at: <http://www.who.int/whr/2002/en/index.html>. Accessed February 2009.
6. Alcohol-attributable deaths and years of potential life lost—United States, 2001. *MMWR Morb Mortal Wkly Rep* 2004; 53: 866–870.
7. Rehm J, Room R, Monteiro M, et al. Alcohol as a risk factor for global burden of disease. *Eur Addict Res* 2003; 9: 157–164.
8. Kim WR, Brown RS Jr, Terrault NA, El-Serag H. Burden of liver disease in the United States: summary of a workshop. *Hepatology* 2002; 36: 227–242.
9. Adachi M, Brenner DA. Clinical syndromes of alcoholic liver disease. *Dig Dis* 2005; 23: 255–263.
10. Leibel W. Epidemiology of alcoholic liver disease. In: Popper H, Schaffner F, eds. *Progress in Liver Disease*. Vol 5. New York, NY: Grune and Stratton; 1976: 494–515.
11. Leibel W. Quantitative aspects of drinking in alcoholic liver cirrhosis. In: Khanna HM, Israel Y, Kalant H, eds. *Alcoholic Liver Pathology*. Toronto, Canada: Toronto Addiction Research Foundation of Ontario; 1975: 1–18.
12. Lindros KO. Alcoholic liver disease: pathobiological aspects. *J Hepatol* 1995; 23(suppl 1): 7–15.
13. Alcoholic liver disease: morphological manifestations. Review by an international group. *Lancet* 1981; 1: 707–711.
14. French SW, Nash J, Shitabata P, et al. Pathology of alcoholic liver disease. VA Cooperative Study Group 119. *Semin Liver Dis* 1993; 13: 154–169.
15. Van Ness MM, Diehl AM. Is liver biopsy useful in the evaluation of patients with chronically elevated liver enzymes? *Ann Intern Med* 1989; 111: 473–478.

16. Mendenhall CL, Seeff L, Diehl AM, et al. Antibodies to hepatitis B virus and hepatitis C virus in alcoholic hepatitis and cirrhosis: their prevalence and clinical relevance. The VA Cooperative Study Group (No. 119). *Hepatology* 1991; 14: 581–589.
17. Mendenhall CL, Moritz T, Rouster S, et al. Epidemiology of hepatitis C among veterans with alcoholic liver disease. The VA Cooperative Study Group 275. *Am J Gastroenterol* 1993; 88: 1022–1026.
18. Orrego H, Blake JE, Blendis LM, Medline A. Prognosis of alcoholic cirrhosis in the presence and absence of alcoholic hepatitis. *Gastroenterology* 1987; 92: 208–214.
19. Levitsky J, Mailliard ME. Diagnosis and therapy of alcoholic liver disease. *Semin Liver Dis* 2004; 24: 233–247.
20. Maddrey WC, Boitnott JK, Bedine MS, et al. Corticosteroid therapy of alcoholic hepatitis. *Gastroenterology* 1978; 75: 193–199.
21. Carithers RL Jr, Herlong HF, Diehl AM, et al. Methylprednisolone therapy in patients with severe alcoholic hepatitis. A randomized multicenter trial. *Ann Intern Med* 1989; 110: 685–690.
22. Ramond MJ, Poynard T, Rueff B, et al. A randomized trial of prednisolone in patients with severe alcoholic hepatitis. *N Engl J Med* 1992; 326: 507–512.
23. Kulkarni K, Tran T, Medrano M, et al. The role of the discriminant factor in the assessment and treatment of alcoholic hepatitis. *J Clin Gastroenterol* 2004; 38: 453–459.
24. Malinchoc M, Kamath PS, Gordon FD, et al. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 2000; 31: 864–871.
25. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001; 33: 464–470.
26. Said A, Williams J, Holden J, et al. Model for End-Stage Liver Disease score predicts mortality across a broad spectrum of liver disease. *J Hepatol* 2004; 40: 897–903.
27. The - MELD model. Available at: <http://www.mayoclinic.org/meld/mayomodel5.html>. Accessed February 2009.
28. Dunn W, Jamil LH, Brown LS, et al. MELD accurately predicts mortality in patients with alcoholic hepatitis. *Hepatology* 2005; 41: 353–358.
29. Srikrueja W, Kyulo NL, Runyon BA, Hu KQ. MELD score is a better prognostic model than Child-Turcotte-Pugh score or discriminant function score in patients with alcoholic hepatitis. *J Hepatol* 2005; 42: 700–706.
30. Sheth M, Riggs M, Patel T. Utility of the Mayo End-Stage Liver Disease (MELD) score in assessing prognosis of patients with alcoholic hepatitis. *BMC Gastroenterol* 2002; 2: 2.
31. Forrest EH, Evans CD, Stewart S, et al. Analysis of factors predictive of mortality in alcoholic hepatitis and derivation and validation of the Glasgow alcoholic hepatitis score. *Gut* 2005; 54: 1174–1179.
32. Lille model. Available at: <http://www.lillemodel.com>. Accessed February 2009.
33. Louvet A, Naveau S, Abdelnour M, et al. The Lille model: a new tool for therapeutic strategy in patients with severe alcoholic hepatitis treated with steroids. *Hepatology* 2007; 45: 1348–1354.
34. Boursier J, Demy M, Moal V, et al. Prediction of outcome in patients with severe acute alcoholic hepatitis treated by corticosteroids. *Hepatology* 2008; 48: S636A.
35. Day CP. Treatment of alcoholic liver disease. *Liver Transpl* 2007; 13: S69–S75.
36. Reynolds TB, Geller HM, Kuzma OT, Redeker AG. Spontaneous decrease in portal pressure with clinical improvement in cirrhosis. *N Engl J Med* 1960; 263: 734–739.
37. Runyon BA. Historical aspects of treatment of patients with cirrhosis and ascites. *Semin Liver Dis* 1997; 17: 163–173.
38. Alexander JF, Lischner MW, Galambos JT. Natural history of alcoholic hepatitis. II. The long-term prognosis. *Am J Gastroenterol* 1971; 56: 515–525.
39. Powell WJ Jr, Klatskin G. Duration of survival in patients with Laennec's cirrhosis. Influence of alcohol withdrawal, and possible effects of recent changes in general management of the disease. *Am J Med* 1968; 44: 406–420.
40. Niemela O, Risteli J, Blake JE, et al. Markers of fibrogenesis and basement membrane formation in alcoholic liver disease. Relation to severity, presence of hepatitis, and alcohol intake. *Gastroenterology* 1990; 98: 1612–1619.
41. Fuller RK, Branchey L, Brightwell DR, et al. Disulfiram treatment of alcoholism. A Veterans Administration cooperative study. *JAMA* 1986; 256: 1449–1455.
42. Kiefer F, Mann K. New achievements and pharmacotherapeutic approaches in the treatment of alcohol dependence. *Eur J Pharmacol* 2005; 526: 163–171.
43. West SL, Garbutt JC, Carey TS, et al. *Pharmacotherapy for Alcohol Dependence*. Rockville, MD: US Department of Health and Human Services; 1999.
44. Snyder JL, Bowers TG. The efficacy of acamprosate and naltrexone in the treatment of alcohol dependence: a relative benefits analysis of randomized controlled trials. *Am J Drug Alcohol Abuse* 2008; 34: 449–461.
45. Anton RF, O'Malley SS, Ciraulo DA, et al. Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. *JAMA* 2006; 295: 2003–2017.
46. Morley KC, Teesson M, Reid SC, et al. Naltrexone versus acamprosate in the treatment of alcohol dependence: a multi-centre, randomized, double-blind, placebo-controlled trial. *Addiction* 2006; 101: 1451–1462.
47. Addolorato G, Leggio L, Ferrulli A, et al. Effectiveness and safety of baclofen for maintenance of alcohol abstinence in alcohol-dependent patients with liver cirrhosis: randomised, double-blind controlled study. *Lancet* 2007; 370: 1915–1922.
48. Garbutt JC, Flannery B. Baclofen for alcoholism. *Lancet* 2007; 370: 1884–1885.
49. Spahr L, Rubbia-Brandt L, Pugin J, et al. Rapid changes in alcoholic hepatitis histology under steroids: correlation with soluble intercellular adhesion molecule-1 in hepatic venous blood. *J Hepatol* 2001; 35: 582–589.
50. Mathurin P, Louvet A, Dharancy S. Treatment of severe forms of alcoholic hepatitis: where are we

- going? *J Gastroenterol Hepatol* 2008; 23(suppl 1): S60–S62.
51. Mendenhall CL, Anderson S, Garcia-Pont P, et al. Short-term and long-term survival in patients with alcoholic hepatitis treated with oxandrolone and prednisolone. *N Engl J Med* 1984; 311: 1464–1470.
 52. Shumaker JB, Resnick RH, Galambos JT, et al. A controlled trial of 6-methylprednisolone in acute alcoholic hepatitis. With a note on published results in encephalopathic patients. *Am J Gastroenterol* 1978; 69: 443–449.
 53. Blitzer BL, Mutchnick MG, Joshi PH, et al. Adrenocorticosteroid therapy in alcoholic hepatitis. A prospective, double-blind randomized study. *Am J Dig Dis* 1977; 22: 477–484.
 54. Theodossi A, Eddleston AL, Williams R. Controlled trial of methylprednisolone therapy in severe acute alcoholic hepatitis. *Gut* 1982; 23: 75–79.
 55. Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials. *Ann Intern Med* 1990; 113: 299–307.
 56. Christensen E, Gluud C. Glucocorticoids are ineffective in alcoholic hepatitis: a meta-analysis adjusting for confounding variables. *Gut* 1995; 37: 113–118.
 57. Mathurin P, Mendenhall CL, Carithers RL Jr, et al. Corticosteroids improve short-term survival in patients with severe alcoholic hepatitis (AH): individual data analysis of the last three randomized placebo controlled double blind trials of corticosteroids in severe AH. *J Hepatol* 2002; 36: 480–487.
 58. Rambaldi A, Saconato HH, Christensen E, et al. Systematic review: glucocorticosteroids for alcoholic hepatitis—a Cochrane Hepato-Biliary Group systematic review with meta-analyses and trial sequential analyses of randomized clinical trials. *Aliment Pharmacol Ther* 2008; 27: 1167–1178.
 59. Mathurin P, O'Grady J, Carithers RL, et al. Corticosteroids improve 28-day survival in patients with severe alcoholic hepatitis: individual data analysis of the last 5 randomised controlled trials. *Hepatology* 2008; 48: S635A.
 60. Rongey C, Kaplowitz N. Current concepts and controversies in the treatment of alcoholic hepatitis. *World J Gastroenterol* 2006; 12: 6909–6921.
 61. Mathurin P, Duchatelle V, Ramond MJ, et al. Survival and prognostic factors in patients with severe alcoholic hepatitis treated with prednisolone. *Gastroenterology* 1996; 110: 1847–1853.
 62. Mathurin P, Abdelnour M, Ramond MJ, et al. Early change in bilirubin levels is an important prognostic factor in severe alcoholic hepatitis treated with prednisolone. *Hepatology* 2003; 38: 1363–1369.
 63. Bird GL, Sheron N, Goka AK, et al. Increased plasma tumor necrosis factor in severe alcoholic hepatitis. *Ann Intern Med* 1990; 112: 917–920.
 64. Felver ME, Mezey E, McGuire M, et al. Plasma tumor necrosis factor alpha predicts decreased long-term survival in severe alcoholic hepatitis. *Alcohol Clin Exp Res* 1990; 14: 255–259.
 65. Shirin H, Bruck R, Aeed H, et al. Pentoxifylline prevents concanavalin A-induced hepatitis by reducing tumor necrosis factor alpha levels and inhibiting adhesion of T lymphocytes to extracellular matrix. *J Hepatol* 1998; 29: 60–67.
 66. McHutchison JG, Runyon BA, Draguescu JO, et al. Pentoxifylline may prevent renal impairment (hepatorenal syndrome) in severe acute alcoholic hepatitis. *Hepatology* 1991; 14: S96A.
 67. Akriviadis E, Botla R, Briggs W, et al. Pentoxifylline improves short-term survival in severe acute alcoholic hepatitis: a double-blind, placebo-controlled trial. *Gastroenterology* 2000; 119: 1637–1648.
 68. Mathurin P. Corticosteroids for alcoholic hepatitis—what's next? *J Hepatol* 2005; 43: 526–533.
 69. Louvet A, Diaz E, Dharancy S, et al. Early switch to pentoxifylline in patients with severe alcoholic hepatitis is inefficient in non-responders to corticosteroids. *J Hepatol* 2008; 48: 465–470.
 70. Calvey H, Davis M, Williams R. Controlled trial of nutritional supplementation, with and without branched chain amino acid enrichment, in treatment of acute alcoholic hepatitis. *J Hepatol* 1985; 1: 141–151.
 71. Lesesne HR, Bozymski EM, Fallon HJ. Treatment of alcoholic hepatitis with encephalopathy. Comparison of prednisolone with caloric supplements. *Gastroenterology* 1978; 74: 169–173.
 72. Nasrallah SM, Galambos JT. Amino acid therapy of alcoholic hepatitis. *Lancet* 1980; 2: 1276–1277.
 73. Galambos JT, Hersch T, Fulenwider JT, et al. Hyperalimentation in alcoholic hepatitis. *Am J Gastroenterol* 1979; 72: 535–541.
 74. Diehl AM, Boitnott JK, Herlong HF, et al. Effect of parenteral amino acid supplementation in alcoholic hepatitis. *Hepatology* 1985; 5: 57–63.
 75. Cabre E, Gonzalez-Huix F, Abad-Lacruz A, et al. Effect of total enteral nutrition on the short-term outcome of severely malnourished cirrhotics. A randomized controlled trial. *Gastroenterology* 1990; 98: 715–720.
 76. Bunout D, Aicardi V, Hirsch S, et al. Nutritional support in hospitalized patients with alcoholic liver disease. *Eur J Clin Nutr* 1989; 43: 615–621.
 77. Nielsen K, Kondrup J, Martinsen L, et al. Long-term oral refeeding of patients with cirrhosis of the liver. *Br J Nutr* 1995; 74: 557–567.
 78. Campillo B, Bories PN, Pornin B, Devanlay M. Influence of liver failure, ascites, and energy expenditure on the response to oral nutrition in alcoholic liver cirrhosis. *Nutrition* 1997; 13: 613–621.
 79. Mendenhall CL, Moritz TE, Roselle GA, et al. A study of oral nutritional support with oxandrolone in malnourished patients with alcoholic hepatitis: results of a Department of Veterans Affairs cooperative study. *Hepatology* 1993; 17: 564–576.
 80. Kearns PJ, Young H, Garcia G, et al. Accelerated improvement of alcoholic liver disease with enteral nutrition. *Gastroenterology* 1992; 102: 200–205.
 81. Cabre E, Rodriguez-Iglesias P, Caballeria J, et al. Short- and long-term outcome of severe alcohol-induced hepatitis treated with steroids or enteral nutrition: a multicenter randomized trial. *Hepatology* 2000; 32: 36–42.
 82. Nompleggi DJ, Bonkovsky HL. Nutritional supplementation in chronic liver disease: an analytical review. *Hepatology* 1994; 19: 518–533.
 83. Mendenhall CL, Anderson S, Weesner RE, et al. Protein-calorie malnutrition associated with alcoholic hepatitis. Veterans Administration Cooperative Study

- Group on Alcoholic Hepatitis. *Am J Med* 1984; 76: 211–222.
84. Patek AJ Jr. Alcohol, malnutrition, and alcoholic cirrhosis. *Am J Clin Nutr* 1979; 32: 1304–1312.
85. Linscheer WG, Patterson JF, Moore EW, et al. Medium and long chain fat absorption in patients with cirrhosis. *J Clin Invest* 1966; 45: 1317–1325.
86. Halsted CH. Nutrition and alcoholic liver disease. *Semin Liver Dis* 2004; 24: 289–304.
87. McCullough AJ, Tavill AS. Disordered energy and protein metabolism in liver disease. *Semin Liver Dis* 1991; 11: 265–277.
88. Spahr L, Rubbia-Brandt L, Frossard JL, et al. Combination of steroids with infliximab or placebo in severe alcoholic hepatitis: a randomized controlled pilot study. *J Hepatol* 2002; 37: 448–455.
89. Tilg H, Jalan R, Kaser A, et al. Anti-tumor necrosis factor- α monoclonal antibody therapy in severe alcoholic hepatitis. *J Hepatol* 2003; 38: 419–425.
90. Mookerjee RP, Tilg H, Williams R, Jalan R. Infliximab and alcoholic hepatitis. *Hepatology* 2004; 40: 499–500; author reply 500–501.
91. Naveau S, Chollet-Martin S, Dharancy S, et al. A double-blind randomized controlled trial of infliximab associated with prednisolone in acute alcoholic hepatitis. *Hepatology* 2004; 39: 1390–1397.
92. Remicade (infliximab) for IV injection. 2005. Available at: <http://www.fda.gov/medwatch/safety/2005/may05.htm>. Accessed February 2009.
93. Menon KV, Stadheim L, Kamath PS, et al. A pilot study of the safety and tolerability of etanercept in patients with alcoholic hepatitis. *Am J Gastroenterol* 2004; 99: 255–260.
94. Boetticher NC, Peine CJ, Kwo P, et al. A randomized, double-blinded, placebo-controlled multicenter trial of etanercept in the treatment of alcoholic hepatitis. *Gastroenterology* 2008; 135: 1953–1960.
95. Phillips M, Curtis H, Portmann B, et al. Antioxidants versus corticosteroids in the treatment of severe alcoholic hepatitis—a randomised clinical trial. *J Hepatol* 2006; 44: 784–790.
96. Stewart S, Prince M, Bassendine M, et al. A randomized trial of antioxidant therapy alone or with corticosteroids in acute alcoholic hepatitis. *J Hepatol* 2007; 47: 277–283.
97. Nguyen-Khac E, Thevenot T, Piquet M, et al. Treatment of severe alcoholic hepatitis with corticosteroids plus *N*-acetylcysteine versus corticoids: planned interim analysis of a multicenter, controlled, randomized trial. *Hepatology* 2008; 48: S634A.
98. Mata JM, Kershenobich D, Villarreal E, Rojkind M. Serum free proline and free hydroxyproline in patients with chronic liver disease. *Gastroenterology* 1975; 68: 1265–1269.
99. Rambaldi A, Gluud C. Colchicine for alcoholic and non-alcoholic liver fibrosis and cirrhosis. *Cochrane Database Syst Rev* 2005; CD002148.
100. Morgan TR, Weiss DG, Nemchauskay B, et al. Colchicine treatment of alcoholic cirrhosis: a randomized, placebo-controlled clinical trial of patient survival. *Gastroenterology* 2005; 128: 882–890.
101. Lieber CS. Role of *S*-adenosyl-*l*-methionine in the treatment of liver diseases. *J Hepatol* 1999; 30: 1155–1159.
102. Rambaldi A, Gluud C. *S*-Adenosyl-*l*-methionine for alcoholic liver diseases. *Cochrane Database Syst Rev* 2006; CD002235.
103. Mato JM, Camara J, Fernandez de Paz J, et al. *S*-Adenosylmethionine in alcoholic liver cirrhosis: a randomized, placebo-controlled, double-blind, multicenter clinical trial. *J Hepatol* 1999; 30: 1081–1089.
104. Israel Y, Kalant H, Orrego H, et al. Experimental alcohol-induced hepatic necrosis: suppression by propylthiouracil. *Proc Natl Acad Sci U S A* 1975; 72: 1137–1141.
105. Rodriguez-Rodriguez E, Gonzalez-Reimers E, Santolaria-Fernandez F, et al. Cytokine levels in acute alcoholic hepatitis: a sequential study. *Drug Alcohol Depend* 1995; 39: 23–27.
106. Rambaldi A, Gluud C. Propylthiouracil for alcoholic liver disease. *Cochrane Database Syst Rev* 2005; CD002800.
107. Mihas AA, Holley P, Koff RS, Hirschowitz BI. Fulminant hepatitis and lymphocyte sensitization due to propylthiouracil. *Gastroenterology* 1976; 70: 770–774.
108. Limaye A, Ruffolo PR. Propylthiouracil-induced fatal hepatic necrosis. *Am J Gastroenterol* 1987; 82: 152–154.
109. Ruiz JK, Rossi GV, Vallejos HA, et al. Fulminant hepatic failure associated with propylthiouracil. *Ann Pharmacother* 2003; 37: 224–228.
110. Deidiker R, de Mello DE. Propylthiouracil-induced fulminant hepatitis: case report and review of the literature. *Pediatr Pathol Lab Med* 1996; 16: 845–852.
111. Rosenak BD, Moser RH, Kilgore B Jr. Treatment of cirrhosis of the liver with testosterone propionate. *Gastroenterology* 1947; 9: 695–704; discussion 749–755.
112. Hall K. Histological changes produced by castration and by sex hormones in the adrenals of normal and of castrated male rats. *Nature* 1937; 140: 318.
113. Rambaldi A, Gluud C. Anabolic-androgenic steroids for alcoholic liver disease. *Cochrane Database Syst Rev* 2006; CD003045.
114. United Network for Organ Sharing. Available at: <http://www.unos.org>. Accessed February 2009.
115. Belle SH, Beringer KC, Detre KM. Liver transplantation for alcoholic liver disease in the United States: 1988 to 1995. *Liver Transpl Surg* 1997; 3: 212–219.
116. Bellamy CO, DiMartini AM, Ruppert K, et al. Liver transplantation for alcoholic cirrhosis: long term follow-up and impact of disease recurrence. *Transplantation* 2001; 72: 619–626.
117. 2006 OPTN/SRTR annual report: transplant data 1996–2005. Available at: http://www.ustransplant.org/annual_reports/archives/2006. Accessed February 2009.
118. Gedaly R, McHugh PP, Johnston TD, et al. Predictors of relapse to alcohol and illicit drugs after liver transplantation for alcoholic liver disease. *Transplantation* 2008; 86: 1090–1095.
119. Moss AH, Siegler M. Should alcoholics compete equally for liver transplantation? *JAMA* 1991; 265: 1295–1298.

120. Gish RG, Lee AH, Keeffe EB, et al. Liver transplantation for patients with alcoholism and end-stage liver disease. *Am J Gastroenterol* 1993; 88: 1337–1342.
121. Berlakovich GA, Steininger R, Herbst F, et al. Efficacy of liver transplantation for alcoholic cirrhosis with respect to recidivism and compliance. *Transplantation* 1994; 58: 560–565.
122. Osorio RW, Ascher NL, Avery M, et al. Predicting recidivism after orthotopic liver transplantation for alcoholic liver disease. *Hepatology* 1994; 20: 105–110.
123. Fabrega E, Crespo J, Casafont F, et al. Alcoholic recidivism after liver transplantation for alcoholic cirrhosis. *J Clin Gastroenterol* 1998; 26: 204–206.
124. Dew MA, DiMartini AF, Steel J, et al. Meta-analysis of risk for relapse to substance use after transplantation of the liver or other solid organs. *Liver Transpl* 2008; 14: 159–172.
125. Miguet M, Monnet E, Vanlemmens C, et al. Predictive factors of alcohol relapse after orthotopic liver transplantation for alcoholic liver disease. *Gastroenterol Clin Biol* 2004; 28: 845–851.
126. Foster PF, Fabrega F, Karademir S, et al. Prediction of abstinence from ethanol in alcoholic recipients following liver transplantation. *Hepatology* 1997; 25: 1469–1477.
127. McCallum S, Masterton G. Liver transplantation for alcoholic liver disease: a systematic review of psychosocial selection criteria. *Alcohol Alcohol* 2006; 41: 358–363.
128. Keeffe EB, Esquivel CO. Controversies in patient selection for liver transplantation. *West J Med* 1993; 159: 586–593.
129. Jauhar S, Talwalkar JA, Schneekloth T, et al. Analysis of factors that predict alcohol relapse following liver transplantation. *Liver Transpl* 2004; 10: 408–411.
130. Pfitzmann R, Schwenzer J, Rayes N, et al. Long-term survival and predictors of relapse after orthotopic liver transplantation for alcoholic liver disease. *Liver Transpl* 2007; 13: 197–205.
131. Mackie J, Groves K, Hoyle A, et al. Orthotopic liver transplantation for alcoholic liver disease: a retrospective analysis of survival, recidivism, and risk factors predisposing to recidivism. *Liver Transpl* 2001; 7: 418–427.
132. DiMartini A, Day N, Dew MA, et al. Alcohol consumption patterns and predictors of use following liver transplantation for alcoholic liver disease. *Liver Transpl* 2006; 12: 813–820.
133. Lim JK, Keeffe EB. Liver transplantation for alcoholic liver disease: current concepts and length of sobriety. *Liver Transpl* 2004; 10: S31–S38.
134. Cuadrado A, Fabrega E, Casafont F, Pons-Romero F. Alcohol recidivism impairs long-term patient survival after orthotopic liver transplantation for alcoholic liver disease. *Liver Transpl* 2005; 11: 420–426.
135. Conjeevaram HS, Hart J, Lissos TW, et al. Rapidly progressive liver injury and fatal alcoholic hepatitis occurring after liver transplantation in alcoholic patients. *Transplantation* 1999; 67: 1562–1568.
136. Tang H, Boulton R, Gunson B, et al. Patterns of alcohol consumption after liver transplantation. *Gut* 1998; 43: 140–145.
137. Dumortier J, Guillaud O, Adham M, et al. Negative impact of de novo malignancies rather than alcohol relapse on survival after liver transplantation for alcoholic cirrhosis: a retrospective analysis of 305 patients in a single center. *Am J Gastroenterol* 2007; 102: 1032–1041.
138. Burra P, Mioni D, Cecchetto A, et al. Histological features after liver transplantation in alcoholic cirrhotics. *J Hepatol* 2001; 34: 716–722.
139. Kelly M, Chick J, Gribble R, et al. Predictors of relapse to harmful alcohol after orthotopic liver transplantation. *Alcohol Alcohol* 2006; 41: 278–283.
140. Gish RG, Lee A, Brooks L, et al. Long-term follow-up of patients diagnosed with alcohol dependence or alcohol abuse who were evaluated for liver transplantation. *Liver Transpl* 2001; 7: 581–587.
141. DiMartini A, Day N, Dew MA, et al. Alcohol use following liver transplantation: a comparison of follow-up methods. *Psychosomatics* 2001; 42: 55–62.
142. Kotlyar DS, Burke A, Campbell MS, Weinrieb RM. A critical review of candidacy for orthotopic liver transplantation in alcoholic liver disease. *Am J Gastroenterol* 2008; 103: 734–743; quiz 744.
143. Everhart JE, Beresford TP. Liver transplantation for alcoholic liver disease: a survey of transplantation programs in the United States. *Liver Transpl Surg* 1997; 3: 220–226.
144. Lucey MR, Brown KA, Everson GT, et al. Minimal criteria for placement of adults on the liver transplant waiting list: a report of a national conference organized by the American Society of Transplant Physicians and the American Association for the Study of Liver Diseases. *Liver Transpl Surg* 1997; 3: 628–637.
145. Veldt BJ, Laine F, Guillygomarch A, et al. Indication of liver transplantation in severe alcoholic liver cirrhosis: quantitative evaluation and optimal timing. *J Hepatol* 2002; 36: 93–98.
146. Lucey MR. Is liver transplantation an appropriate treatment for acute alcoholic hepatitis? *J Hepatol* 2002; 36: 829–831.
147. Neuberger J, Schulz KH, Day C, et al. Transplantation for alcoholic liver disease. *J Hepatol* 2002; 36: 130–137.
148. Pageaux GP, Bismuth M, Perney P, et al. Alcohol relapse after liver transplantation for alcoholic liver disease: does it matter? *J Hepatol* 2003; 38: 629–634.
149. Wells JT, Said A, Agni R, et al. The impact of acute alcoholic hepatitis in the explanted recipient liver on outcome after liver transplantation. *Liver Transpl* 2007; 13: 1728–1735.
150. Vanlemmens C, Di Martino V, Milan C, et al. Immediate listing for liver transplantation versus standard care for Child-Pugh stage B alcoholic cirrhosis: a randomized trial. *Ann Intern Med* 2009; 150: 153–161.
151. Duvoux C, Delacroix I, Richardet JP, et al. Increased incidence of oropharyngeal squamous cell carcinomas after liver transplantation for alcoholic cirrhosis. *Transplantation* 1999; 67: 418–421.
152. Presser SJ, Schumacher G, Neuhaus R, et al. De novo esophageal neoplasia after liver transplantation. *Liver Transpl* 2007; 13: 443–450.
153. Pageaux GP, Michel J, Coste V, et al. Alcoholic cirrhosis is a good indication for liver transplantation, even for cases of recidivism. *Gut* 1999; 45: 421–426.
154. Coffman KL, Hoffman A, Sher L, et al. Treatment of the postoperative alcoholic liver transplant recipient

- with other addictions. *Liver Transpl Surg* 1997; 3: 322–327.
155. Lucey MR, Merion RM, Henley KS, et al. Selection for and outcome of liver transplantation in alcoholic liver disease. *Gastroenterology* 1992; 102: 1736–1741.
156. Tripp LE, Clemons JR, Goldstein RR, McCarthy Stewart L. Drinking patterns in liver transplant recipients. *Psychosomatics* 1996; 37: 249–253.
157. Burra P, Mioni D, Cillo U, et al. Long-term medical and psycho-social evaluation of patients undergoing orthotopic liver transplantation for alcoholic liver disease. *Transpl Int* 2000; 13(suppl 1): S174–S178.
158. Bjornsson E, Olsson J, Rydell A, et al. Long-term follow-up of patients with alcoholic liver disease after liver transplantation in Sweden: impact of structured management on recidivism. *Scand J Gastroenterol* 2005; 40: 206–216.
159. Bravata DM, Olkin I, Barnato AE, et al. Employment and alcohol use after liver transplantation for alcoholic and nonalcoholic liver disease: a systematic review. *Liver Transpl* 2001; 7: 191–203.
160. Jain A, DiMartini A, Kashyap R, et al. Long-term follow-up after liver transplantation for alcoholic liver disease under tacrolimus. *Transplantation* 2000; 70: 1335–1342.
161. Pereira SP, Howard LM, Muiesan P, et al. Quality of life after liver transplantation for alcoholic liver disease. *Liver Transpl* 2000; 6: 762–768.
162. Platz KP, Mueller AR, Spree E, et al. Liver transplantation for alcoholic cirrhosis. *Transpl Int* 2000; 13(suppl 1): S127–S130.
163. DiMartini A, Jain A, Irish W, et al. Outcome of liver transplantation in critically ill patients with alcoholic cirrhosis: survival according to medical variables and sobriety. *Transplantation* 1998; 66: 298–302.
164. Gerhardt TC, Goldstein RM, Urschel HC, et al. Alcohol use following liver transplantation for alcoholic cirrhosis. *Transplantation* 1996; 62: 1060–1063.
165. Tringali RA, Trzepacz PT, DiMartini A, Dew MA. Assessment and follow-up of alcohol-dependent liver transplantation patients. A clinical cohort. *Gen Hosp Psychiatry* 1996; 18: 708–775.
166. Doffoel M, Fratte S, Vanlemmens C, et al. Results of liver transplantation in 75 French patients with alcoholic cirrhosis. Comparison with a non-alcoholic cirrhotic group [abstract]. *Hepatology* 1992; 16: 50A.
167. Kumar S, Stauber RE, Gavalier JS, et al. Orthotopic liver transplantation for alcoholic liver disease. *Hepatology* 1990; 11: 159–164.

SPECIAL FEATURE

Treatment of Primary Vertebral Tumors

Harshpal Singh, MD, Scott A. Meyer, MD, and Arthur L. Jenkins III, MD, FACS

¹Mount Sinai School of Medicine, New York, NY

OUTLINE

BENIGN TUMORS

Bony Tumors
Cartilaginous Tumors
Vascular Tumors
Plasma Cell Dyscrasias
Embryonic Rests

MALIGNANT TUMORS

Bony Tumors
Vascular Tumors
Embryonic Rests
Plasma Cell Dyscrasias

CONCLUSION

types. *Mt Sinai J Med* 76:499–504, 2009. © 2009 Mount Sinai School of Medicine

Key Words: spine tumor treatment, vertebral tumors.

Primary vertebral tumors encompass a wide variety of tumor types classified by their cell of origin: bony tumors, cartilaginous tumors, vascular tumors, plasma cell dyscrasias, and tumors that arise from embryonic rests. The tumors are further divided into malignant or benign subtypes on the basis of their clinical progression, histopathological evidence of invasiveness, and response to therapy. We will provide a brief overview of the treatment of primary vertebral tumors.

ABSTRACT

Advances in the treatment of disease and detection through advanced imaging have led to an increase in the discovery of vertebral tumors. Although the majority of symptomatic spinal lesions are metastatic in origin, a significant number of them arise primarily in the spine. These lesions encompass a wide variety of tumor types classified by their cell of origin: bony tumors, cartilaginous tumors, vascular tumors, plasma cell dyscrasias, and tumors that arise from embryonic rests. Further classification of these tumors into malignant or benign subtypes is based on their clinical progression, histopathological evidence of invasiveness, and response to therapy. We provide a brief overview and description of primary tumors as well as treatment paradigms for the individual tumor

BENIGN TUMORS

Bony Tumors

Osteoid osteomas and osteblastomas are osteoblast derivatives that typically occur in long bones of the body. However, approximately 20% of osteoid osteomas and 40% of osteblastomas occur in the spine.^{1,2} Both produce osteoid and woven bone, but they are differentiated by their size; osteblastomas tend to be larger (>1.5 cm) and more aggressive and can undergo malignant transformation. Osteomas are small, benign, and self-limited lesions that have little potential for growth.

Osteoid osteomas are anywhere from 4 to 10 times more common than osteblastomas and are commonly found in the thoracolumbar spine.³ They predominantly occur in the posterior elements, such as the transverse process, facet, lamina, or pedicle. There is a male predominance ranging anywhere from 2:1 to 4:1.^{4,5} Patients commonly present during adolescence with complaints of localized neck or back pain that worsens with weight-bearing activity. Pain is typically nocturnal and can be alleviated with aspirin or other nonsteroidal anti-inflammatory drugs. On computed tomography (CT) imaging, these

Address Correspondence to:

Arthur L. Jenkins III
Mount Sinai School of Medicine
New York, NY
Email: arthur.jenkins@mountsinai.org

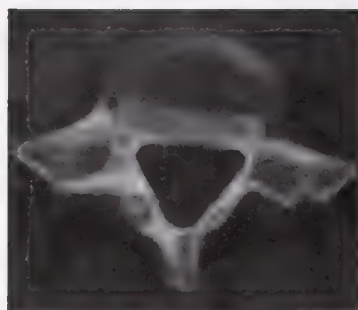


Fig 1. Axial noncontrast computed tomography scan demonstrating an osteoid osteoma of the right L5 lamina.

tumors have an osteosclerotic rim with a central calcific nidus (Figure 1). They have low intensity on T1- and T2-weighted magnetic resonance imaging (MRI). However, the most sensitive imaging modality for these tumors is the technetium bone scan, with only a single reported false negative.^{6,7} Surgical resection involves exposure and curettage of the lesion. During surgery, the nidus appears as a punctiform spot with a reddish color, contrasting with the surrounding white bone. The appropriate technique is to expose the lesion and curette the nidus by the gradual excision of minimal reactive bone. These lesions are curative with a gross total resection.

Osteoblastomas make up less than 1% of primary bone tumors and also occur predominantly in the posterior elements of the spine. Epidural and soft tissue extension is frequently observed. Anywhere from 25% to 70% of patients with an osteoblastoma present with a neurological deficit versus 0% to 30% of patients who have an osteoid osteoma.⁷ They are greater than 1.5 cm in size and tend to be lucent and lytic in appearance. The technetium scan remains the most sensitive imaging modality. The surgical appearance and excision techniques are also similar to those for osteoid osteomas. Aggressive surgical management is necessary because these tumors are radioresistant; recurrence may occur as long as 9 years after surgery. Additionally, malignant transformation into an osteosarcoma is possible.⁸

Cartilaginous Tumors

Chondromas are benign rests of cartilage within cancellous bone that result from the failed migration of chondrocytes, with less than 4% originating in the spine.⁹ They are labeled according to the site of origin; those arising from the medullary cavity are termed *enchondromas*, whereas those arising from the cortical surface are called *periosteal chondromas*. Enchondromas are expansile, and periosteal chondromas are exophytic. Chondromas are slow-growing, isolated lesions that present with local

tenderness and may cause a radiculopathy or myelopathy by direct neural compression. Multiple lesions are suggestive of a multiple chondromatosis syndrome, such as Ollier disease or Maffucci syndrome.¹⁰ A CT scan demonstrates a radiolucent, erosive lesion sometimes with stippled patterns of calcification. Although these lesions can be observed, surgical excision is recommended with histological examination of the entire specimen, as sarcomatous cells may be present in only a fraction of the tumor. The surgical technique involves curettage of the entire lesion with a graft and instrumentation if necessary.

Vascular Tumors

Hemangiomas have an incidence of up to 11% in autopsy series. Although these lesions are typically asymptomatic, they can present with localized or diffuse pain. Myelopathy and radiculopathy can result from neural compression by tumors. A CT scan demonstrates thickened vertical trabeculations surrounding dilated vascular spaces, giving a characteristic honeycomb appearance¹¹ (Figure 2). Treatment of symptomatic lesions is accomplished through surgical resection with or without postoperative radiotherapy. Vertebroplasty can be used for managing pain in those with an intraosseous lesion; however, the posterior bone margin must be intact, or the injected material may extravasate into the spinal canal and cause cord compression. Preoperative angiography with embolization prior to resection is a powerful adjunct in minimizing operative blood loss. Radiation therapy of 3000 to 4000 rad has also been used in treatment.

Angiolipomas are composed of mature fat cells and abnormal vascular elements.¹² There are only

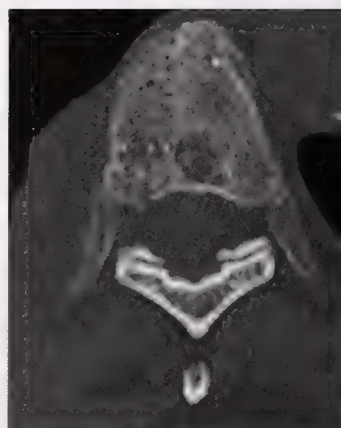


Fig 2. Axial noncontrast computed tomography scan demonstrating the characteristic honeycomb appearance of a vertebral hemangioma.

about 100 cases reported in the literature with an incidence of 0.14% to 1.2% of spinal axis tumors.¹³ Presenting symptoms are those of spinal cord compression. Angiolipomas may be hypodense or isodense on CT imaging. MRI is considered the gold standard; the lesions are hypointense to fat on T1 and demonstrate inhomogeneous enhancement on postcontrast sequences. These tumors are curable after surgical resection.

Aneurysmal bone cysts are blood-filled cavities separated by septae of trabecular bone or fibrous tissue containing osteoclastic giant cells.¹⁴ They represent approximately 15% of primary spine tumors and are more common in the lumbar spine.¹⁵ Patients most commonly present when they are 20 years old or younger¹⁶ with complaints of localized pain. CT imaging demonstrates involvement of the posterior elements and associated vertebral body; a balloon-like expansile soft tissue mass surrounded by a shell-like cortical margin is typical. Fine needle biopsy tends to be nondiagnostic, so open biopsy with curettage resection is recommended. Preoperative embolization minimizes operative blood loss. The majority of recurrences are seen less than 6 months after surgical resection, and repeat radical resection followed by postoperative radiotherapy is recommended.

Plasma Cell Dyscrasias

Eosinophilic granulomas are benign, self-limited non-neoplastic collections of lipid-laden phagocytic histiocytes of the reticuloendothelial system. They are part of a more generalized disorder called Langerhans cell histiocytosis. About 10% to 15% occur in the spine, with patients presenting with neurological deficits from direct neural compression or constitutional symptoms such as fever and weight loss. Radiographically, oval radiolucent areas of bone destruction with peripheral sclerosis are seen. Vertebral collapse producing flat, planelike vertebrae called *vertebra plana* is a classic radiographic finding.^{17,18} When this is seen, a skeletal survey is indicated to discover multifocal lesions. The syndrome is characterized by spontaneous regression and occasional reactivation, so treatment is conservative in nature. Modalities include partial lesion curettage, chemotherapy, and low-dose radiation.¹¹ Rest and external mobilization remain the most effective therapy, unless cord compression is seen.

Embryonic Rests

Giant cell tumors are slow-growing, locally invasive tumors that typically occur in the sacrum but can

metastasize. The tumors are composed of multinucleated, osteoclastic giant cells that are thought to arise from mononuclear cells of macrophage origin. They are primarily in the vertebral body and occur primarily in females in their twenties and thirties.¹⁹ The most common presenting complaint is back pain, but neurological deficits occur by direct neural compression. Because the sacrum is a typical site, sacral nerve root involvement may result in bowel or bladder dysfunction. A CT scan shows a lytic, expansile lesion with a soap bubble-like vertebral expansion (Figure 3). MRI better demonstrates the bony and soft tissue component of the lesion. The ideal treatment for giant cell tumors is complete resection, but en bloc resection increases the potential morbidity associated with surgery. Extensive surgical resection may result in spinal instability, requiring instrumentation and fusion. Radiation therapy is often used as postoperative adjuvant therapy in incomplete resections with recommended dosing of 25 to 45 Gy with standard fractionated therapy.²⁰ There is concern about sarcomatous degeneration after radiation therapy, with one study showing an 11% incidence.²¹ Stereotactic radiosurgery has also been used for spinal tumors and has been reported for giant cell tumors as well. Necrosis and hemorrhage are not uncommon.

MALIGNANT TUMORS

Bony Tumors

Osteosarcoma is the most common sarcoma of the spine, accounting for 3% to 15% of primary spine tumors.²² Most tumors arise in the posterior elements, with the thoracic spine and sacral spine being the most common sites of occurrence.²³ Patients typically

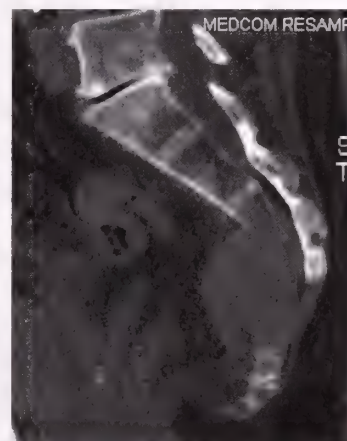


Fig 3. Sagittal noncontrast computed tomography scan demonstrating a giant cell tumor of the sacrococcygeal region.

present with complaints of axial back pain or radicular pain, but weakness is seen in more than half. In the setting of Paget's disease, malignant degeneration to osteosarcoma occurs in less than 1%.²⁴ CT imaging can show osteosclerotic or osteoblastic features. MRI shows mixed intensity signals, but heterogeneous enhancement on postcontrast sequences helps distinguish tumor from muscle. When it is metastatic, the survival rate drops dramatically, with 11% survival at 20 months. Although the prognosis is generally poor, treatment involves aggressive surgical resection followed by chemotherapy because these tumors are relatively radioresistant. Combined anterior and posterior surgical resection is typical.

Chondrosarcomas are malignant hyaline-secreting tumors that account for 25% of primary malignant spine neoplasms.⁸ Most tumors occur on the lateral side of the vertebral body near the costotransverse junction and are felt to originate from the cartilaginous proximal rib end. Patients present with complaints of focal back pain, but neurological deficit is seen in over half of patients. These deficits include radiculopathy, myelopathy, and cauda equina syndrome. Radiographic imaging shows a lytic lesion with a calcified soft tissue mass. CT and MRI are recommended to help delineate bony and soft tissue involvement. Treatment involves radical surgical resection because these tumors are resistant to chemotherapy and radiation.⁸ However, total excision is difficult, and recurrence or metastasis is common. Preoperative angiography and embolization are recommended to decrease intraoperative blood loss.

Vascular Tumors

Malignant vascular tumors include hemangiopericytomas, hemangioendotheliomas, and angiosarcomas. All 3 types are lytic lesions associated with cortical destruction. CT and MRI demonstrate vascular channels at the periphery of the lesions. Angiography shows well-demarcated areas of staining and early draining veins. Hemangiopericytomas additionally demonstrate a vascular pedicle that gives off feeders circling the lesion.

There are only a few dozen case reports of spinal hemangiopericytomas in the literature. Histologically, slit-like vascular spaces are found that, when large, may assume a classic staghorn appearance. Radiation therapy has been proved to be of no benefit, and the mainstay of treatment is surgical resection. The reported recurrence rate is 76% to 80%, and the metastatic rate ranges from 23% to 64%.²⁵

Hemangioendotheliomas are derived from the vascular endothelium and rarely originate in the

spine. The tumor cells show endothelial differentiation characterized by intracytoplasmic inclusions seen as vacuoles and containing erythrocytes.²⁶ The prognosis of the tumor depends on its histopathological features and location.²⁷

Angiosarcomas are derived malignant endothelial cells lining anastomotic vascular channels. All age groups are affected, but there is an almost 2:1 male preponderance. On presentation, patients complain of dull, aching pain over the affected region. The prognosis of skeletal angiosarcoma is poor; the majority of patients die within 1 year of diagnosis.

Treatment for all 3 types of malignant vascular-derived tumors focuses on complete surgical resection of the lesion. Preoperative angiography and tumor embolization are recommended to reduce surgical blood loss. Biopsy is not recommended because of the high incidence of bleeding complications. Radiation and chemotherapy have been advocated for the more malignant lesions, but their efficacy remains unproved.

Embryonic Rests

Chordomas are primary malignant tumors that arise from the embryonic remnants of the notochord and occur in less than 1 in 1,000,000.²⁸ They are considered slow-growing sarcomas associated with significant local bone destruction and neural compression, with the sacrococcygeal region most commonly affected. Patients typically present with local pain, cord compression, or nerve root compression. A CT scan demonstrates osteolytic, osteosclerotic, or mixed areas of bone destruction (Figure 4). Gadolinium contrast-enhanced MRI allows for better delineation of these tumors. Because these lesions show a poor response to ionizing radiation and chemotherapy, current treatment strategies involve radical surgical resection when possible followed by a high-energy photon or proton beam for recurrent/residual disease. En bloc excision is difficult because of anatomic limitations. These tumors have a high recurrence rate, with local recurrence seen as soon as 2 to 4 years after the initial treatment.

Ewing's sarcoma is a primitive neuroectodermal tumor that usually originates from the medullary cavity of long bones. About 20% of patients present with gross metastatic disease. Presenting symptoms include localized pain with point tenderness. Chromosomal studies reveal a unique reciprocal translocation, t(11;22)(q24;12). Radiographs reveal a classic onion peel periosteal reaction. CT and MRI can demonstrate the soft tissue paraspinal mass associated with the bony lesion. The initial treatment of Ewing's sarcoma is systemic chemotherapy with



Fig 4. Sagittal noncontrast computed tomography scan demonstrating a replacement of the sacrococcygeal region by chordoma.

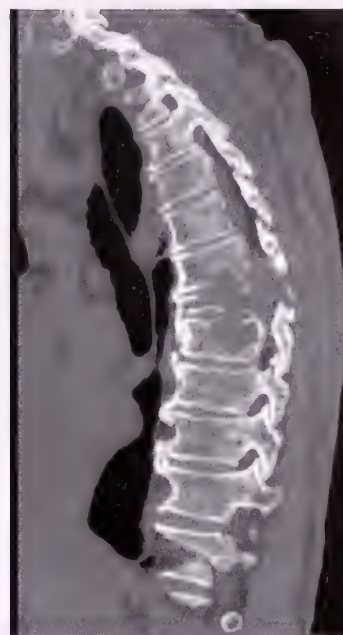


Fig 5. Sagittal noncontrast computed tomography of the thoracic spine demonstrating a plasmacytoma with epidural extension into the spinal canal causing myelopathy.

high-dose intermittent agents including vincristine, actinomycin, and cyclophosphamide. An alternative treatment is bone marrow transplantation combined with radiation and chemotherapy. Surgical resection can be considered to reduce pain, decompress neural elements, and achieve spinal stability. Radiation therapy has also been used as adjunctive therapy; classic dose recommendations include 40 to 45 Gy to the whole field followed by a booster or 50 to 60 Gy with a coned-down field.¹¹

Plasma Cell Dyscrasias

Multiple myeloma and solitary plasmacytoma represent a spectrum of clonal B cell tumors involving bone. Multiple myeloma involves the malignant transformation of plasma cells in bone marrow that subsequently produces a monoclonal gammopathy. Back pain is the usual initial complaint, but a neurological deficit is usually what causes patients to seek medical attention. Radiographically, compression fractures, lytic lesions, and osteopenia are seen. The classic punched-out lesions of multiple myeloma are multifocal, well-circumscribed hypodense lesions with sclerotic margins. The diffuse marrow involvement is best appreciated when the noncontrast sequences are compared to the contrast-enhanced sequences on MRI. Serum and urine electrophoresis demonstrates monoclonal hypergammaglobulinemia or Bence-Jones protein, respectively. Treatment involves chemotherapy and/or radiation when there

is no significant cord compromise or spinal instability; the usual agents include cyclophosphamide, melphalan, carmustine, lomustine, and concurrent steroids. Surgical decompression and possible stabilization are pursued in cases of spinal instability. A halo vest can be used in cases of cervical myeloma during treatment with irradiation; this prevents neurological deterioration from instability and may promote stable bony reconstitution. Percutaneous vertebral augmentation (kyphoplasty or vertebroplasty) has shown significant improvement for ongoing pain.²⁹ The overall prognosis is poor, with a median survival of 4 years.³⁰ Poorer survival is associated with epidural compression, anemia, hypercalcemia, renal failure, and hyperproteinuria.

Solitary plasmacytoma is defined as a localized plasma cell malignancy that has a male predominance.³¹ Twenty-five to fifty percent of all lesions occur in the spine. Patients present with localized pain or a neurological deficit from neural compression. Radiographically, the lesion appears as a large, well-defined cyst-like lesion. CT imaging shows a lytic lesion with sclerotic margins (Figure 5). A bone scan and skeletal survey are recommended to rule out additional lesions, and bone marrow biopsy is recommended to rule out multiple myeloma. Unfortunately, half of all patients progress to multiple myeloma within 5 years of diagnosis. The treatment is radiotherapy for patients that present without rapid neurological deterioration. Local recurrence is rare

when doses greater than 45 Gy are given. Local control rates of up to 96% and survival up to 11 years have been reported. Surgical resection is indicated when neural compression and significant neurological deterioration occur. A complete resection and cure can be achieved, usually involving reconstruction and surgical stabilization. In cases of subtotal resection, radiation and chemotherapy are recommended. Paraprotein levels are followed for evidence of recurrence or dissemination.

Multiple types of primary vertebral tumors have been discussed. The treatment of each type of tumor is dependent on its degree of invasiveness, growth, malignant potential, and susceptibility to medical therapy.

DISCLOSURES

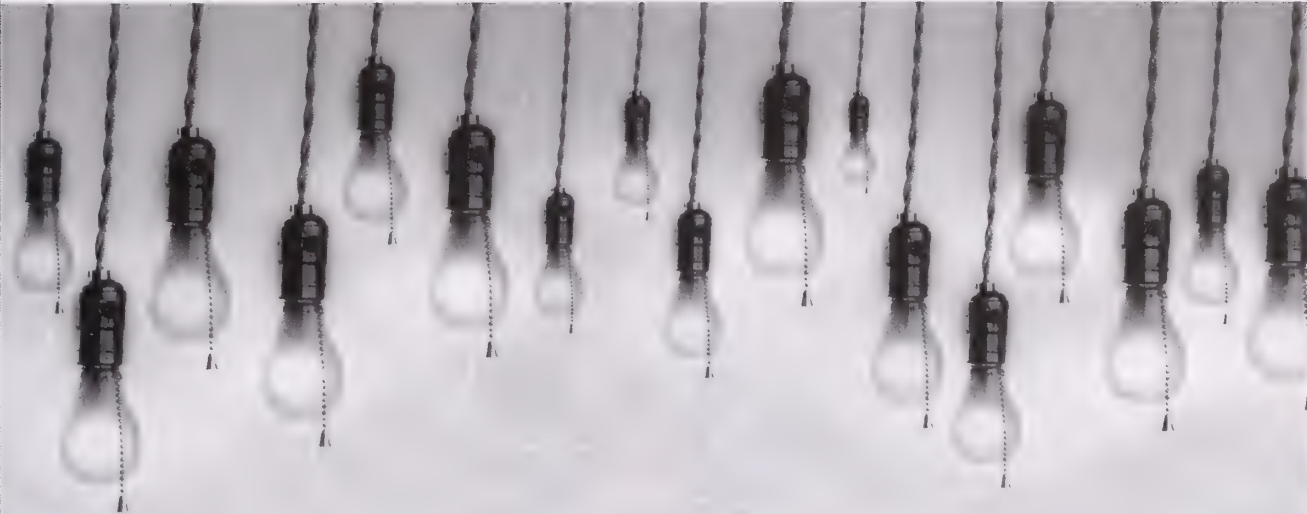
Potential conflict of interest: Nothing to report.

REFERENCES

1. Jackson RP, Reckling FW, Mants FA. Osteoid osteoma and osteoblastoma. Similar histologic lesions with different natural histories. *Clin Orthop Relat Res* 1977; 128: 303–313.
2. Marsh BW, Bonfiglio M, Brady LP. Benign osteoblastoma: range of manifestations. *J Bone Joint Surg Am* 1975; 57: 1–9.
3. Azouz EM, Kozlowski K, Marton D. Osteoid osteoma and osteoblastoma of the spine in children. Report of 22 cases with brief literature review. *Pediatr Radiol* 1986; 16: 25–31.
4. Nemoto O, Moser RP Jr, Van Dam BE. Osteoblastoma of the spine. A review of 75 cases. *Spine* 1990; 15: 1272–1280.
5. Pettine KA, Klassen RA. Osteoid-osteoma and osteoblastoma of the spine. *J Bone Joint Surg Am* 1986; 68: 354–361.
6. Fehring TK, Green NE. Negative radionuclide scan in osteoid osteoma. A case report. *Clin Orthop Relat Res* 1984; 185: 245–249.
7. Kan P, Schmidt MH. Osteoid osteoma and osteoblastoma of the spine. *Neurosurg Clin N Am* 2008; 19: 65–70.
8. McLoughlin GS, Sciubba DM, Wolinsky J-P. Chondroma/chondrosarcoma of the spine. *Neurosurg Clin N Am* 2008; 19: 57–63.
9. Mordard M, De Tribolet N, Janzer R. Chondromas of the spine: report of two cases and review of the literature. *Br J Neurosurg* 1993; 7: 551–556.
10. Silve C, Jüppner H. Ollier disease. *Orphanet J Rare Dis* 2006; 1: 804–812.
11. Sundaresan N, Boriani S, Rothman A, Holtzman R. Tumors of the osseous spine. *J Neurooncol* 2004; 69: 273–290.
12. Lin JJ, Lin F. Two entities in angiolipoma. A study of 459 cases of lipoma with review of literature on infiltrating angiolipoma. *Cancer* 1974; 34: 720–727.
13. Hungs M, Pare LS. Spinal angiolipoma: case report and literature review. *J Spinal Cord Med* 2008; 31: 315–318.
14. Aho HJ, Aho AJ, Einola S. Aneurysmal bone cyst, a study of ultrastructure and malignant transformation. *Virchows Arch A Pathol Anat Histol* 1982; 395: 169–179.
15. Burch S, Hu S, Berven S. Aneurysmal bone cysts of the spine. *Neurosurg Clin N Am* 2008; 19: 41–47.
16. Leithner A, Windhager R, Lang S, et al. Aneurysmal bone cyst. A population-based epidemiologic study and literature review. *Clin Orthop Relat Res* 1999; 363: 176–179.
17. Azouz EM, Saigal G, Rodriguez MM, Podda A. Eosinophilic granuloma of the cervical spine. *Pediatr Radiol* 2005; 35: 103–115.
18. Buchmann J. Osteochondritis of the vertebral body. *J Bone Joint Surg* 1927; 9: 55–66.
19. Luther N, Bilsky MH, Härtl R. Giant cell tumor of the spine. *Neurosurg Clin N Am* 2008; 19: 49–55.
20. Caudell JJ, Ballo MT, Zagars GK, et al. Radiotherapy in the management of giant cell tumor of bone. *Int J Radiat Oncol Biol Phys* 2003; 57: 158–165.
21. Leggon RE, Zlotecki R, Reith J, Scarborough MT. Giant cell tumor of the pelvis and sacrum: 17 cases and analysis of the literature. *Clin Orthop Relat Res* 2004; 423: 196–207.
22. Wang VY, Potts M, Chou D. Sarcoma and the spinal column. *Neurosurg Clin N Am* 2008; 19: 71–80.
23. Ozaki T, Flege S, Liljenqvist U, et al. Osteosarcoma of the spine: experience of the Cooperative Osteosarcoma Study Group. *Cancer* 2002; 94: 1069–1077.
24. Sharma H, Mehdi SA, MacDuff E, et al. Paget sarcoma of the spine: Scottish Bone Tumor Registry experience. *Spine* 2006; 31: 1344–1350.
25. Ijiri K, Yuasa S, Yone K, et al. Primary epidural hemangiopericytoma in the lumbar spine. *Spine* 2002; 27: 189–192.
26. Gokhan GA, Akyuz M, Gurer IE, Tuncer R. Epithelioid hemangioendothelioma derived from the spine region: case report and review of the literature. *Wien Klin Wochenschr* 2006; 118: 358–361.
27. Faust J, Schmidt M, Eysel P, Wanitschke R. Epithelioid hemangioendothelioma of the spine. *Med Klin (Munich)* 2001; 96: 740–744.
28. McMaster ML, Goldstein AM, Bromley CM, et al. Chordoma: incidence and survival patterns in the United States, 1973–1995. *Cancer Causes Control* 2001; 12: 1–11.
29. Dudeney S, Lieberman IH, Reinhardt MK, Hussein M. Kyphoplasty in the treatment of osteolytic vertebral compression fractures as a result of multiple myeloma. *J Clin Oncol* 2002; 20: 2382–2387.
30. Bilsky MH, Azeem S. Multiple myeloma: primary bone tumor with systemic manifestations. *Neurosurg Clin N Am* 2008; 19: 31–40.
31. Voulgaris S, Partheni M, Gousias K, et al. Solitary plasmacytoma of the upper cervical spine: therapeutic considerations. *J Neurosurg Sci* 2008; 52: 55–59.



don't publish online in the dark...



let our open linking network
shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-

profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at www.crossref.org, and make sure all your electronically published work is CrossRef-enabled.



CROSSREF.ORG | THE CITATION LINKING BACKBONE

200 WHEELER ROAD, 6TH FLOOR, BURLINGTON, MA 01803 • 781-359-2442



Discover
Something
Great

Distinctive Content

Content

Journals • Backfile Collections
Reference Works & Encyclopedias
OnlineBooks™
Current Protocols Lab Manuals
Evidence-Based Resources

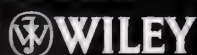
Access

Single article/chapter or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage

Capabilities

E-mail Alerts • EarlyView®
Citation download & tracking
Remote access • Mobile Edition®

Promote the process of scientific discovery with a different breed of research system: **Wiley InterScience®**. Long distinguished by its scope and extensive coverage of emerging disciplines, Wiley's authoritative content spans the spectrum of scientific, medical, technical, and business specialties. The dynamic **Wiley InterScience** platform adds value to your research by providing simple desktop access to this preeminent digital library of journals, reference works, books, laboratory manuals, evidence-based databases, and more. Researchers reap the combined benefits of top-tier resources and up-to-the-minute functionality. Experience the difference **Wiley InterScience** can make at your institution today.



1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

MOUNT SINAI JOURNAL OF MEDICINE

VOL. 75, NO. 6

NOVEMBER/DECEMBER 2009

Issue Theme

Advances in Obstetrics
and Gynecology:
A Brave New World

Theme Editor

Rhoda Sperling, MD

A Journal of Translational and Personalized Medicine



Assisted Reproductive Technology
Non-invasive Prenatal Diagnosis
Achieving a Safety Culture in Obstetrics
Obesity and Pregnancy
Treatment of Uterine Leiomyomata

KLF6 and KLF6-SV1 in Ovarian Cancer
Surgical Management of Cervical Cancer
Chemotherapy of Gynecologic Disorders
Bloodless Surgery in Gynecologic Oncology

Special Features

Breast Magnetic Resonance Imaging: An Overview for Non-radiologists
Genomics of Common Diseases: Cardiovascular Disease
Treatment of Metastatic Spinal Tumors

 **WILEY-BLACKWELL**
Publishers since 1807

**Newly
Relaunched!**



**Keep Abreast of the Latest Trends in
Translational and Personalized Medicine...
Enter Your Subscription Today !**

MOUNT SINAI JOURNAL OF MEDICINE

A Journal of Translational and Personalized Medicine

Editor-in-Chief:

Penny A. Asbell, MD, FACS, MBA

Mount Sinai School of Medicine, New York, NY

2008 Themes and Guest Editors

Quality of Healthcare and Racial
Health Disparities (Mark Chassin)

9/11 World Trade Center Disaster:
Past and Future (Robin Herbert,
Jacqueline Moline)

Anxiety and Depression:
- Leading Edge of Therapy
(Dennis Charney, Mathew Sanjay)

Diabetes Type 1:
New Light on Etiology,
New Approaches to Management
(Derek Leroith)

Hospital Medicine:
An Ascending New Field
(Andrew Dunn)

Advanced Lung Disease:
Meeting the Challenges
(Maria Padilla)

Continuously published since 1934 and included in *Index Medicus* from the very first volume, the **Mount Sinai Journal of Medicine (MSJM)** has a distinguished career as a general medical journal for a worldwide audience.

Building upon this tradition of education across medical specialties, **MSJM** was re-launched in 2007 as a journal of translational and personalized medicine, two disciplines in the forefront of modern healthcare. **MSJM** features articles on the evolving nature of clinical care, its close linkage to bio-medical research and the resulting opportunities for enhanced patient care.

Dedicated to improving communication among all participants in medicine today, **MSJM** is an excellent source of up-to-date medical information for healthcare clinicians, educators, researchers and policy makers worldwide. **MSJM** will inspire the creative thinking needed to create new ways to understand, prevent, and treat human disease.

Each issue includes several quality, in-depth review articles related to a single theme and several others on subjects of diverse clinical interest. All articles are written by renowned authorities in their fields, from both Mount Sinai and elsewhere, under the supervision of Guest Editors.

 **WILEY-BLACKWELL**

Enter Your 2009 Subscription Today!



Your 2009 subscription will include the six 2009 print issues of the **Mount Sinai Journal of Medicine** (Volume 76), plus concurrent full-text online access to these six issues as they are posted on Wiley's Internet publishing platform.

Two Ways for Individuals to Subscribe:

1. Mount Sinai Alumna/Alumnus:
\$40.00 (Worldwide)
2. Personal Subscription:
\$95.00 (U.S./Canada/Mexico)
\$120.00 (Outside North America)

Complimentary Online Access for Institutions!

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

PERSONAL SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 personal subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$95.00 (U.S./Canada/Mexico) ☐ \$120.00 (Outside North America)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

Method of Payment (Canada residents please add GST. Personal rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **John Wiley & Sons, Inc.** Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA

Card # _____ Exp. Date _____

Signature _____
(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM0909

MSJM MEMBER SUBSCRIPTION ORDER FORM

☐ **Yes**, please enter my 2009 Alumna/Alumnus Subscription to the **Mount Sinai Journal of Medicine** at the following subscription rate.

☐ \$40.00 (Worldwide)

Please note that all subscriptions are entered on a calendar year basis, regardless of the month that the subscription form is sent back. You will receive all issues from January 2009 through December 2009.

METHOD OF PAYMENT (Alumna/Alumnus rate subscriptions must be paid by personal check or credit card. They may not be resold or used as library copies.)

☐ Check/Money Order (Payable to **The Mount Sinai School of Medicine**. Payment must be made in U.S. dollars and drawn on a U.S. bank.)

☐ Charge My Credit Card: ☐ American Express ☐ MasterCard ☐ VISA ☐ Other

Card # _____ Exp. Date _____

Signature _____
(credit card orders not valid without signature)

Ship To:

Name _____

Address _____

City _____ State _____ Zip _____

MSJM09

COMPLIMENTARY ONLINE ACCESS FOR INSTITUTIONS

Mount Sinai Journal of Medicine

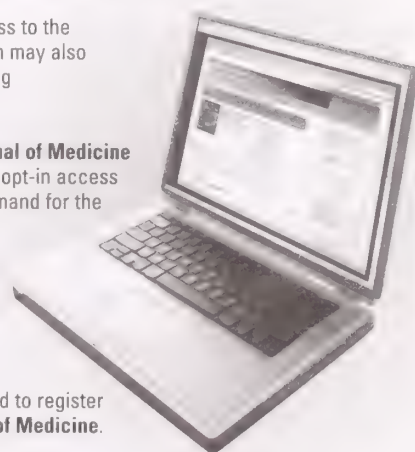
A Journal of Translational and Personalized Medicine

2009. Volume 76. 6 issues per year. Print ISSN: 0027-2507. Online ISSN: 1931-7581.

Your institution can opt to receive complimentary online access to the **Mount Sinai Journal of Medicine** during 2009. Your institution may also request one complimentary print subscription when registering for complimentary online access.

Institutional online and print pricing for the **Mount Sinai Journal of Medicine** will be set in 2009, subsequent to introductory complimentary opt-in access during 2008 and 2009. Pricing will be based on the overall demand for the Journal and on the size and type of the subscribing institution (e.g. academic, government, hospital, or corporate). Institutions that receive complimentary online access during 2008 and 2009 will have archival access in subsequent years according to Wiley's licensing terms and conditions.

Ask your librarian to e-mail optinaccess@wiley.co.uk or visit www.interscience.wiley.com/msjregister for more details and to register for complimentary online access to the **Mount Sinai Journal of Medicine**.





NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS MAIL

PERMIT NO. 2277

HOBOKEN, NJ

POSTAGE WILL BE PAID BY ADDRESSEE

JOHN WILEY & SONS INC
ATTN: PETER JANSIC, MAIL STOP 8-02
111 RIVER STREET
HOBOKEN, NJ 07030-9441



PLACE
POSTAGE
STAMP
HERE

INES SANTOS
MOUNT SINAI JOURNAL OF MEDICINE
ONE GUSTAV L. LEVY PLACE BOX 1094
NEW YORK, NY 10029-6574

**Library
Routing
Card**

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

Editor-in-Chief

Penny A. Asbell

Deputy Editor

David H. Adams

Associate Editors

Stuart Aaronson
Erwin Bottinger
Dennis S. Charney
Valentin Fuster
Wayne Goodman
Jonathan Halperin
Ravi Iyengar
Mary Klotman
Paul Klotman
Ihor R. Lemischka
Michael L. Marin
Thomas G. McGinn
Ira Nash
Maria New
Rosamond Rhodes
Hugh Sampson
Mary Sano
Jeffrey H. Silverstein
Albert Siu
Anthony Squire
Fred Suchy
Savio Woo

Editorial Board

William J. Bremner	University of Washington
Thomas M. Coffman	Duke University
Dwight Evans	University of Pennsylvania
Gabriel Haddad	University of California, San Diego
Heidi E. Hamm	Vanderbilt University
Matti Häyry	University of Manchester
Eve J. Higginbotham	Morehouse School of Medicine
Katherine A. High	University of Pennsylvania
Tony Hope	University of Oxford
John H. Krystal	Yale University
Raju Kucherlapati	Harvard Medical School
Robert J. Lefkowitz	Duke University
Robert Lisak	Wayne State University
George Lister	University of Texas
Philip A. Marsden	University of Toronto
Charles Nemeroff	Emory University
Scott Rauch	Harvard Medical School
Bernard Roizman	University of Chicago
Alan F. Schatzberg	Stanford University
Inder Verma	The Salk Institute
Jeffrey A. Whitsett	University of Cincinnati
Allan Wolkoff	Albert Einstein College of Medicine
Noam Zohar	Bar-Ilan University

Theme Editor

Rhoda Sperling, MD
Research Director and Vice Chair
Department of Obstetrics, Gynecology and Reproductive Science
Mount Sinai School of Medicine, New York, NY

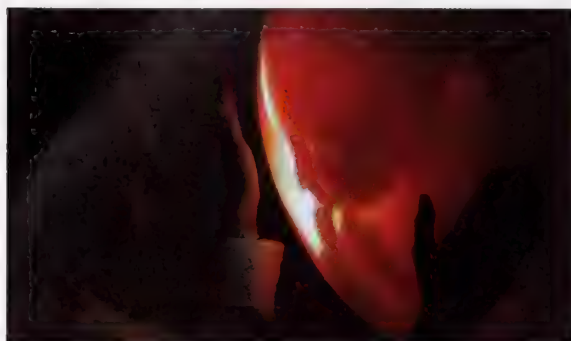
Editorial Director

Stephen Cohen
stephen.cohen@mssm.edu
212-241-6304

Publisher

John Wiley & Sons
Anna Wesley, Associate Editor
awesley@wiley.com

ADVANCES IN OBSTETRICS AND GYNECOLOGY: A BRAVE NEW WORLD



A bubble-like application of 4D ultrasound. 2D ultrasound produces a flat image; 3D ultrasound adds width, height and depth; 4D ultrasound adds movement. With 4D ultrasound, clinicians and parents alike can experience a fetus in all its spatial and temporal reality.

Reproduced with permission of Royal Philips Electronics.

Future *MSJM* Themes and Theme Editors

- Alzheimer's disease (Nikos Robakis)
- Endovascular surgery (Michael Marin)
- Genes, race, ethnicity and the environment (Gary Butts)
- Drug discovery in academic settings (Ravi Iyengar)
- Obesity (Derek LeRoith)
- Computer-assisted radiology (Burton Drayor)
- Translational approaches to cancer treatment (Steven Burakoff)
- Children's health and the environment (Philip Landrigan)

©2009 Mount Sinai School of Medicine. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. Phone: (978) 750-8400; Fax: (978) 750-4770. Such requests and other permission inquiries should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; Phone: (201) 748-6011; Fax: (201) 748-6008; E-mail: permreq@wiley.com.

Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine (Print ISSN: 0027-2507; Online ISSN: 1931-7581 at Wiley InterScience, www.interscience.wiley.com) is published twelve times per year for the Mount Sinai School of Medicine by Wiley Subscription Services, a Wiley Company, 111 River Street, Hoboken, NJ 07030.

Send subscription inquiries to: Subscription Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-6645.

Offprint sales and inquiries should be addressed to the Reprint Billing Department, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030; (201) 748-8776. All other inquiries should be addressed to the Customer Service Department at (201) 748-6645.

Individual subscription price: \$95.00 in North America, \$120.00 outside North America. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been delivered. For more subscription prices and information, please visit www.mountsinaijournal.org.

Change of address: Please forward six weeks prior to move: present mailing label with change of address. Please enclose a copy of the mailing label or cite your subscriber reference number in order to expedite handling. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of: *Mount Sinai Journal of Medicine*, c/o John Wiley & Sons, Inc., Attn: Journals Admin Dept. UK, 111 River Street, Hoboken, NJ 07030.

Cancellations: Subscription cancellations will not be accepted after the first issue has been mailed. Periodicals postage paid at Hoboken, NJ and at additional mailing offices.

Postmaster: Send address changes to *Mount Sinai Journal of Medicine*, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Printed in the United States of America.

Advertising inquiries should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030. Phone: (201) 748-6921; Fax: (201) 748-6207. Advertising in this journal is published with a caveat emptor understanding. The Publisher, Editors, and members of the Editorial Board do not imply warranty of products, nor quality, validity, or approval of education material offered by such advertising.

Indexed by: Biological Abstracts® (Thomson ISI), Chemical Abstracts Service/SciFinder (ACS), Current Contents®/Clinical Medicine (Thomson ISI), EMBASE/Excerpta Medica (Elsevier), Index Medicus/MEDLINE/PubMed (NLM), Science Citation Index Expanded™ (Thomson ISI), SCOPUS (Elsevier), Web of Science® (Thomson ISI).

This journal is printed on acid-free paper.

Disclaimer: All articles published, including but not limited to original research, reviews, clinical notes, clinical pearls, prismatic cases, clinician's corner, editorials, reports, and letters, represent the opinions and views of the authors and do not reflect any official policy or medical opinion of the Mount Sinai School of Medicine, the institutions with which the authors are affiliated, or the Publisher unless this is clearly specified. Articles published herein should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by physicians for any particular patient or disorder. Although the Editors and Publisher believe that drug selections and dosages and the specifications and usage of equipment and devices as set forth herein are in accord with current recommendations and practice at the time of publication, they accept no legal responsibility for any errors or omissions, and make no warranty, express or implied, with respect to material contained herein.

MOUNT SINAI JOURNAL OF MEDICINE

A JOURNAL OF TRANSLATIONAL AND PERSONALIZED MEDICINE

CONTENTS

VOL 76, NO 6

Theme Introduction

- 505** Advances in Obstetrics and Gynecology: A Brave New World
Rhoda Sperling

Theme Articles

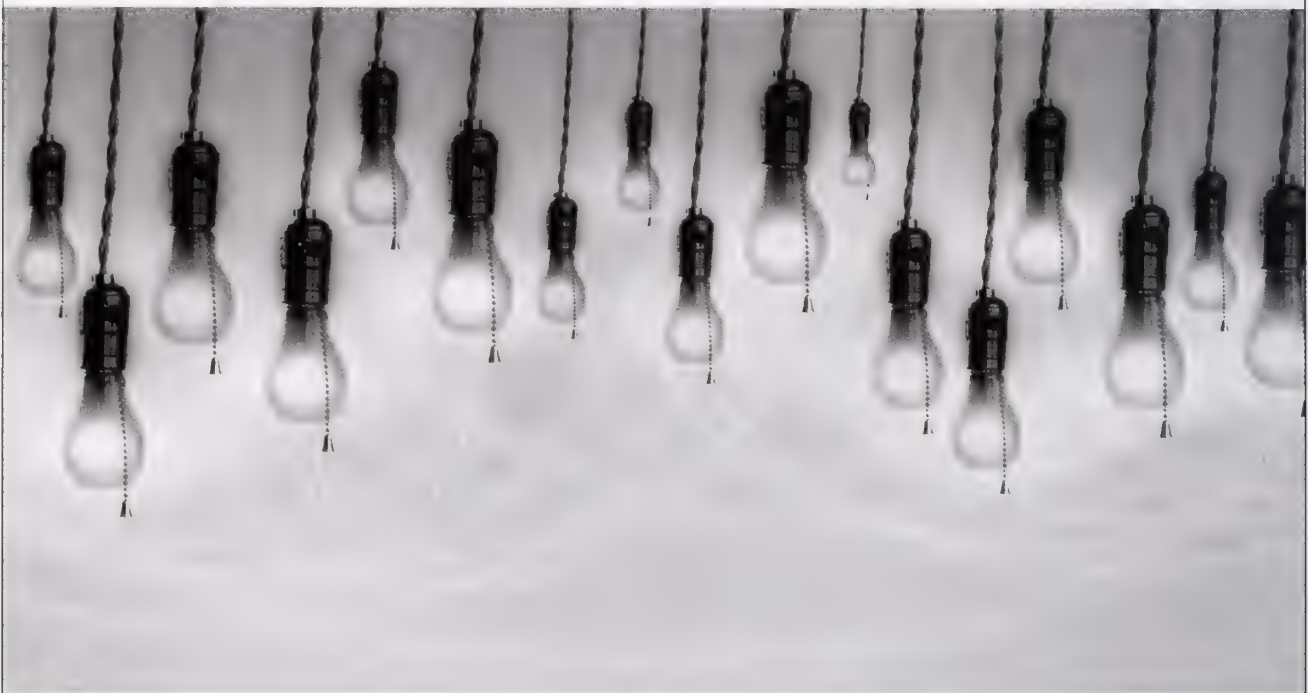
- 506** Advances and Controversies in Assisted Reproductive Technology
Gerardo Vela, Martha Luna, Benjamin Sandler, Alan H. Copperman
- 521** Non-invasive Prenatal Diagnosis: Past, Present and Future
Christian Litton, Joanne Stone, Keith Eddleman, Men-Jean Lee
- 529** Achieving a Safety Culture in Obstetrics
Erin DuPree, Loraine O'Neill, Rebecca M. Anderson
- 539** Obesity and Pregnancy: Implications and Management Strategies for Providers
Taraneh Shirazian, Sreekala Raghavan
- 546** A Comprehensive Approach to the Treatment of Uterine Leiomyomata
Charles Ascher-Walsh
- 557** Emerging Roles of KLF6 and KLF6-SC1 in Ovarian Cancer Progression and Treatment
Analisa DiFeo, Goutham Narla, John A. Martignetti
- 567** Recent Advances in the Surgical Management of Cervical Cancer
Konstantin Zakashansky, William H. Bradley, Linus Chuang, Jamal Rahaman, Peter Bottino
- 577** Chemotherapy for Gynecologic Cancers
Jamal Rahaman
- 589** Bloodless Surgery in Gynecologic Oncology
Nimesh P. Nagsarsheth, Fahimeh Sasan

Special Features

- 598** Breast Magnetic Resonance Imaging: An Overview for Non-radiologists
Laurie Margolies, Richard Ha
- 606** Current Surgical Management of Ulcerative Colitis
Alexis Grucela, Randolph M. Steinhagen
- 613** The Genomics of Cardiovascular Disease
Leslie Manace, Tapap Godiwala, Mark Babyatsky



don't publish online in the dark...



let our open linking network
shed some light.

This journal is CrossRef-ed! As a scientist or scholar navigating digital content, you've come to expect dynamic linking from citation to original source. As an author, you expect ready dissemination of your work to others. CrossRef is a collaborative, cross-publisher reference linking network that allows users to click on a citation and be taken directly to the full-text location. An independent, non-

profit organization that employs open standards, CrossRef was established in cooperation with leading research publishers and the International DOI Foundation. Since the service went live in 2000, CrossRef has linked millions of articles from thousands of journals. Find out more at www.crossref.org, and make sure all your electronically published work is CrossRef-enabled.



CROSSREF.ORG | THE CITATION LINKING BACKBONE

200 WHEELER ROAD, 6TH FLOOR, BURLINGTON, MA 01803 • 781-359-2442

THEME INTRODUCTION

Advances in Obstetrics and Gynecology: A Brave New World

Rhoda Sperling, MD

Department of Obstetrics, Gynecology, and Reproductive Science, Mount Sinai School of Medicine,
New York, NY

This special edition of the *Mount Sinai Journal of Medicine* is devoted to advances in the practice of obstetrics and gynecology. The diversity of the topics that are presented is a reflection of the scientific breadth of a discipline that provides healthcare for women throughout their reproductive lifespan. Elucidating the molecular basis of disease pathogenesis, learning how to use new technologies, incorporating evidence-based findings into practice guidelines, and creating a culture of patient safety are the basic challenges that must be met as we continue to advance the healthcare of women.

I would like to thank the journal editors for allowing the Department of Obstetrics, Gynecology, and Reproductive Science to showcase its work and collaborations in this theme issue. I would also like to thank Dr. Men-Jean Lee and Dr. Konstantin Zakhshansky for their editorial assistance with manuscript preparation as well as Ms. Ashaki Smith for her administrative support.

DISCLOSURES

Potential conflict of interest: Nothing to report.

Address Correspondence to:

Rhoda Sperling

Department of Obstetrics,
Gynecology, and Reproductive
Science
Mount Sinai School of Medicine
New York, NY
Email: rhoda.sperling@mssm.edu

Advances and Controversies in Assisted Reproductive Technology

Gerardo Vela, MD, Martha Luna, MD, Benjamin Sandler, MD,
and Alan B. Copperman, MD

Mount Sinai School of Medicine, New York, NY

OUTLINE

INTRODUCTION

PATIENT SELECTION AND STIMULATION

Indications for In Vitro Fertilization
Preimplantation Genetic Diagnosis and Screening
Fertility Preservation
Evaluating Ovarian Reserve
Ovarian Stimulation

OOCYTE RETRIEVAL AND FERTILIZATION

Oocyte Retrieval
Fertilization Methods
In Vitro Maturation
Gamete and Embryo Cryopreservation

EMBRYO SELECTION AND TRANSFER

IN VITRO FERTILIZATION OUTCOMES

CONCLUSION

ABSTRACT

As the incidence and awareness of infertility have increased in the last decades, reproductive endocrinologists and scientists have striven to improve the therapeutic options being offered to patients. One of the most advanced and efficient technologies currently being offered is assisted reproductive technology. Among the various techniques that are comprised by assisted reproductive technology, in vitro fertilization is the most widely

studied and used, being responsible for approximately 1% of all live births in the United States. As this technology has evolved, many controversies have arisen, and it is the purpose of this article to review what in the authors' opinion are the most current and controversial aspects of the whole process of in vitro fertilization and its related techniques. *Mt Sinai J Med* 76:506–520, 2009. © 2009 Mount Sinai School of Medicine

Key Words: assisted reproductive technology, fertility preservation, in vitro fertilization, infertility, preimplantation genetic diagnosis, review.

INTRODUCTION

As the incidence and awareness of infertility have increased in the last decades, reproductive endocrinologists and scientists have striven to improve the therapeutic options being offered to patients. It was not long ago that laparotomy-requiring microsurgical techniques were replaced by a less invasive laparoscopic approach. In the same way, laparoscopic procedures are now being replaced by even

By optimizing the maturity, quality, and quantity of gametes and allowing them to fertilize in a controlled environment, we are now able to assist and confirm fertilization, and we are learning to diagnose and select healthier embryos so that we can keep improving pregnancy rates and outcomes while reducing the number of multiple and genetically abnormal pregnancies.

Address Correspondence to:

Alan B. Copperman
Mount Sinai School of Medicine
New York, NY
Reproductive Medicine
Associates of New York
635 Madison Ave, 10th Floor
New York, NY 10022
Email: acopperman@rmany.com

less invasive, more efficient assisted reproduction techniques.^{1,2} Because of significant advances in pharmacology, cellular biology, and genetics, more control is being achieved over the entire process of conception. By optimizing the maturity, quality, and quantity of gametes and allowing them to fertilize in a controlled environment, we are now able to assist and confirm fertilization, and we are learning to diagnose and select healthier embryos so that we can keep improving pregnancy rates and outcomes while reducing the number of multiple and genetically abnormal pregnancies. *Assisted reproductive technology* (ART) is a term used when we refer to the manipulation and handling of gametes with the objective of achieving a pregnancy. According to the Centers for Disease Control and Prevention,³ ART includes all fertility treatments in which oocytes, sperm, and embryos are handled. In general, ART procedures involve surgically aspirating oocytes after controlled ovarian hyperstimulation (COH), inseminating them in the laboratory, and transferring the fertilized oocytes to the uterus. Treatments in which only sperm are handled [ie, intrauterine insemination (IUI)] or procedures in which patients undergo ovulation induction without the intention of having oocytes retrieved are not classified as ART. This technology includes, but is not limited to, in vitro fertilization (IVF), gamete intrafallopian transfer (GIFT), zygote intrafallopian transfer (ZIFT), embryo cryopreservation, oocyte or embryo donation, and gestational surrogacy. Assisted reproductive techniques have led to approximately 3 million births up to 2006, as reported by the European Society of Human Reproduction and Embryology,⁴ and it is estimated that approximately 1% of all live births in the United States are conceived by this technology.⁵ Since the

It is estimated that approximately 1% of all live births in the United States are conceived by this technology.

birth of Louis Brown, which was reported in 1978 by Steptoe and Edwards,⁶ much of the research and many of the advances in ART have focused on IVF. Other techniques, including zygote intrafallopian transfer and gamete intrafallopian transfer, were introduced in the mid 1980s, but they have since fallen out of favor. As IVF success rates have improved in the past decades, these techniques have been largely abandoned because they require a more invasive/more expensive procedure and do not possess a clear benefit over IVF with respect to success rates.⁷⁻¹⁰ According to the Society for

Assisted Reproductive Technology, 99.7% of fresh nondonor oocyte or embryo ART procedures in the United States are IVF.⁵

For the purpose of this review, we will, therefore, be focusing primarily on IVF and the techniques that surround it. In an effort to provide a more defined and clearer approach, we have divided IVF chronologically into 3 stages:

- Patient selection and stimulation.
- Oocyte retrieval and fertilization.
- Embryo selection, transfer, and outcome.

Each of these stages will be discussed independently, with an emphasis on the current clinical controversies surrounding them.

PATIENT SELECTION AND STIMULATION

Indications for In Vitro Fertilization

Even though IVF was originally created to overcome fallopian tube obstruction,⁶ the indications have expanded. In fact, as the efficiency of IVF keeps increasing, it is becoming accepted as a therapeutic alternative for many infertility-causing diseases and fertility-related issues (Table 1). In most of the world, IVF is now commonly accepted as the most effective and one of the most commonly used treatment modalities for infertility.

Tubal Factor Infertility.

When it comes to irreparable tubal damage, there is little controversy that IVF should be considered the treatment of choice. Specifically, IVF should be considered to be first-line treatment in patients with tubal blockage, severe pelvic adhesions, or a history of bilateral salpingectomy or previously failed salpingoplasty. In patients who have partially damaged

Table 1. Current Indications for In Vitro Fertilization.

Indication	Current Status
Tubal factor	Severe: absolute Mild to moderate: relative
Male factor	Severe: absolute Mild to moderate: relative
Endometriosis	Relative
Anovulation	Relative
Immunological	Relative
Unexplained	Relative
Preimplantation genetic diagnosis and screening	Absolute
Fertility preservation	Absolute for oocyte and embryo cryopreservation

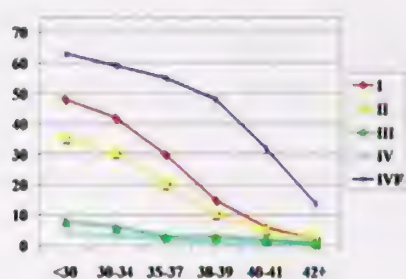


Fig 1. Pregnancy rates after IUF and after surgical repair of different stages of tubal disease. Data from Gomel²⁰ and pregnancy rates of RMA of NY between 2002-2006 (unpublished data).

tubes or who have been surgically sterilized, many aspects have to be considered before a specific line of treatment is chosen: surgery or IVF.¹¹ Distal tubal disease has been treated surgically (via laparotomy or laparoscopy), with success rates ranging from 10% to 35%; unfortunately, at a rate of 5% to 20%, these conceptions implant in the damaged fallopian tube and result in ectopic pregnancies.¹²⁻¹⁵ When the obstruction is located in the proximal portion of the fallopian tube, other treatment options exist. Tubal cannulation^{16,17} or microsurgery^{18,19} can be offered to these patients; however, the success rates are not as high as those obtained with IVF. These alternative treatment options should be considered in patients with ethical and religious concerns. If both proximal and distal disease exists, then IVF is clearly the primary treatment. When we counsel our patients about surgical treatment, it is our responsibility to inform them about the disadvantages of this modality: delay of pregnancy, increased risk of ectopic pregnancies, potential risk of further adhesion formation, risk of general anesthesia, and risk of infection, bleeding, and internal organ damage. Clearly, the higher the success rates are in any given IVF center, the better an option IVF is over tubal reconstructive surgery (Figure 1).

Hydrosalpinx management in IVF patients has also been somewhat controversial. Accumulated evidence has demonstrated that patients with a communicating hydrosalpinx have a lower rate of embryonic implantation after embryo transfer. Data from our group have demonstrated that hydrosalpinx fluid is embryotoxic²¹ and increases the inflammatory milieu of the endometrium.²² It is now clear that when hydrosalpinges are documented on a sonogram or hysterosalpingography, patient counseling should include the risks and benefits of performing either prophylactic salpingectomy or surgical isolation of the damaged fallopian tube from the endometrial cavity by cauterization of the interstitial portion of

the tube (to prevent hydrosalpingeal fluid reflux) prior to embryo transfer.²³⁻²⁵

Endometriosis.

The relationship between infertility and endometriosis has long been enigmatic but has recently begun to be clarified. The prevalence of endometriosis in infertile women ranges from 25% to 50% versus 5% in the fertile population.^{26,27} The definitive diagnosis of endometriosis includes the direct visualization and pathological confirmation of endometriotic lesions. Although the diagnosis requires laparoscopy, the sonographic finding of an endometriotic ovarian collection yields a high suspicion index for the disease. Mild and severe endometriosis has been treated laparoscopically with cumulative pregnancy rates as high as 70%.^{28,29} However, when these results are translated to monthly fecundity rates, they are not higher than 6% per month. During the last several years, the role of diagnostic laparoscopy has been debated,³⁰ and the number of patients undergoing a laparoscopic evaluation as part of their initial workup has decreased.³¹ Subsequently, the number of patients treated surgically for endometriosis-related infertility has been decreasing. The rationale behind treating endometriosis with IVF stems from the idea of bypassing the anatomic alterations and the hostile peritoneal environment and optimizing the likelihood of success by increasing the number of oocytes and selecting the embryos to be transferred.³² Currently, the use of IVF as the treatment for endometriosis-associated infertility is recommended in patients who have a concomitant male factor or severe tubal disease or for whom surgical treatment has previously failed.³³ The American Society of Reproductive Medicine recommends IVF when surgical treatment for moderate and severe endometriosis has not been successful.³⁴ In our opinion, before a treatment option is chosen, the male status, tubal patency, severity of the disease, duration of infertility, patient age, ovarian reserve status, previous treatments, and symptomatology must all be considered.

Male Factor.

A male factor is the sole cause of infertility in approximately 20% of couples undergoing treatment for infertility, but it is a contributing factor in at least 30% to 40% of cases.³⁵ When an abnormal semen analysis is identified, it is important for the male to be evaluated by a qualified urologist. Depending on the etiology and severity of the male factor, a variety of treatment options can be offered in an effort to improve the quantity or quality of spermatozoa.

If no treatable cause is identified or if previous treatments have failed, then IUI or IVF should be considered. In the presence of a mild to moderate male factor (ie, at least >5,000,000 motile sperm), IUI in combination with COH in young women yields a pregnancy rate of approximately 12% per cycle.³⁶ It is important to consider variables such as the wife's age and other potential contributory factors when the recommendation of treatment modality is made.

In case of severe oligospermia or astenospermia, IVF with intracytoplasmic sperm injection (ICSI) may be necessary to achieve fertilization and pregnancy. ICSI, a procedure in which a single spermatozoid is injected into the oocyte by micromanipulation, has pregnancy rates comparable to those obtained by conventional insemination IVF in patients without an associated male factor. In patients with azospermia, important advances in the field of urology/andrology allow us to retrieve spermatozooids directly from the epididymus or the testicle in azospermic patients, and this gives them the opportunity of conceiving with their own sperm through ICSI.

Anovulation.

Anovulatory disorders such as hypothalamic hypogonadism and polycystic ovarian syndrome (PCOS) are often managed effectively with ovulation-inducing medications. When these treatments fail, IVF could be considered.³⁷ Ovarian stimulation is particularly complex in this group of patients because they can easily go from being unresponsive to stimulation to suffering from severe ovarian hyperstimulation syndrome (OHSS). Some authors have suggested that the concomitant use of metformin while patients are being stimulated for IVF not only might increase the pregnancy rates but also could decrease the incidence of OHSS.³⁸ Conversely, a recent meta-analysis by Costello *et al.*³⁹ did not demonstrate any benefits from the cotreatment with metformin.

Another advantage of IVF in patients with anovulatory infertility is the prevention of high-order multiple pregnancies. Patients who undergo ovulation induction and demonstrate a high ovarian response are counseled about cycle cancellation to avoid this outcome. Although some authors have reported favorable results by converting these cycles to IVF,⁴⁰ in our experience (Grinfeld L, unpublished data), it may be reasonable to consider IVF rather than COH as a first-line treatment for patients with PCOS, as it improves pregnancy rates and results in a significantly lower high-order multiple-pregnancy rate.

Immunological Infertility.

The presence of antisperm antibodies (ASAs) in women's serum, cervical mucus, and sperm has been reported by some to be associated with infertility. It is postulated that these antibodies affect conception by altering sperm motility, cervical mucous interaction, and even oocyte fertilization in both in vivo and in vitro settings.⁴¹ Different therapeutic strategies can be considered when patients with ASAs are being treated. Glucocorticoids are no longer used to decrease ASA production as there can be severe associated side effects. Alternative strategies have been used, including the removal of antibodies bound to sperm (with an in vitro sperm wash or capacitation) and the facilitation of the fertilization process by direct insemination of the eggs or injection of the sperm into the oocyte with ICSI.⁴² The pregnancy rates in the former treatments are somewhat diminished, whereas the use of ICSI offers patients success rates as high as those of non-ASA-affected couples undergoing conventional IVF.⁴³

Unexplained Infertility.

The diagnosis of unexplained infertility is made when a couple undergoes a standard evaluation for infertility and no clear cause is identified. The incidence has been reported to be as high as 30% of the infertile population.⁴⁴ Many therapeutic options have been reported and studied. There is some controversy about the potential advantages of IVF over simply observation for couples experiencing idiopathic infertility. The Centers for Disease Control and Prevention reported a live birth rate of 30.4% in 2003, which is comparable to the rates of other infertile patients and clearly higher than the rate with attempted natural conception.⁴⁵ However, a meta-analysis reported by Pandian *et al.*⁴⁶ in 2002 and reviewed in 2005 did not demonstrate an advantage of IVF over expectant management, clomiphene citrate, IUI with or without ovarian stimulation, or gamete intrafallopian transfer in terms of live birth rates. As success rates improve for IVF, the balance may swing in favor of this technique for the treatment of unexplained infertility.

Cost should also be considered as a variable when the appropriate treatment strategy for idiopathic infertility is being determined. According to Karande *et al.*,⁴⁷ the cost of a single pregnancy in a group in which IVF is the first line of treatment was \$57,161 versus \$20,019 in a standard treatment group that included IVF as the ultimate treatment. Conversely, Reindollar *et al.*⁴⁸ showed that taking a fast track to IVF was actually the most cost-effective of approaches.

Other Indications. Utilization of IVF and ART is no longer limited to infertile patients. New

Utilization of IVF and ART is no longer limited to infertile patients.

technologies are now available that allow the assessment of the genetic or chromosomal composition of embryos, giving couples the option to select embryos before transferring them to the uterus. This gives a great advantage to patients that have an increased risk of passing a disease to their children and may also convey an advantage to those with an increased risk of aneuploidy. Another relatively new indication for IVF and ART is fertility preservation. Cryopreservation of embryos, sperm, and even oocytes is becoming a feasible option that can now be offered to patients who might need to either postpone or protect their ability to procreate because of social or medical reasons.

Preimplantation Genetic Diagnosis and Screening

Preimplantation genetic diagnosis (PGD) and preimplantation genetic screening (PGS) are procedures in which embryos obtained from IVF are biopsied in order to obtain specific genetic and chromosomal information by different techniques. PGD was first described in 1990 by Handyside *et al.*⁴⁹ after their team was able to transfer only female embryos in 2 women known to be carriers of X-linked diseases. The usual technique for PGD includes the biopsy of 1 or, in some cases, 2 blastomeres on 3-day-old embryos, and either fluorescence in situ hybridization (FISH) or polymerase chain reaction is used to identify specific genetic or chromosomal traits. After the results are obtained, unaffected embryos are selected and transferred.

When PGD is done exclusively to avoid the transfer of aneuploid embryos [by identifying certain chromosomes with FISH or via newer technologies such as comparative genomic hybridization (CGH) and single nucleotide polymorphism analysis on a microarray chip], it can be more specifically defined as PGS.⁵⁰ The higher incidence of aneuploid embryos in older patients has been clearly demonstrated.^{51,52} Regardless of the morphological appearance, embryos derived from the oocytes of older women demonstrate a lower implantation rate and a higher miscarriage rate after IVF treatments.⁵³ Theoretically, if we could identify euploid embryos and transfer only those, the implantation rates should increase, and the miscarriage rates should decrease. Many investigators have attempted to

achieve this goal, but to date, these studies either have design flaws or have been contradictory in their results.^{51,52,54,55} A recent meta-analysis by Twisk *et al.*,⁵⁶ which included 2 randomized controlled trials, did not find a statistically significant difference in the clinical pregnancy rate (15% in PGS versus 20% in controls) or live birth rate (11% in PGS versus 15% in controls). Mastenbroek *et al.*⁵⁷ published in 2007 a large, multicenter, randomized, double-blinded trial demonstrating that PGS with a 1-cell biopsy using FISH analysis in women between 35 to 41 years not only did not improve but instead significantly reduced the ongoing pregnancy and live birth rates after IVF. Some authors have questioned these results and claimed that the lower pregnancy rates in the PGS group could be due to the transfer of embryos with inconclusive biopsy results or due to poor embryo micromanipulation. Certainly, this points out the need for further studies in order to reach definitive conclusions about PGS and advanced maternal age.

Chromosomal abnormalities are responsible for at least 65% of first-trimester miscarriages.⁵⁸ The same is true for patients with a history of unexplained recurrent miscarriages.^{59,60} Hypothetically, if we select chromosomally normal embryos for transfer, the miscarriage rate should be reduced. Some authors have published data suggesting that aneuploidy screening in these embryos significantly reduces the pregnancy loss rate.^{61,62} The adequacy of the control groups in these studies has been questioned, and some authors have even reported contradictory results.⁶⁰ Larger, multicenter, and better controlled trials are required before routine PGS can be recommended in this group of patients.

Patients with recurrent implantation failure,^{63,64} severe male factor infertility,^{65,66} and chromosomal translocations⁶⁷ have a higher incidence of aneuploidy and, therefore, could theoretically benefit from the use of PGS. Unfortunately, once again, data are either limited or controversial, and no general conclusions can be made about the use of PGS in these patients.

With respect to PGS, the American Society of Reproductive Medicine states that patients must fully understand the limitations, risk of error, and lack of evidence of improvement in live birth rates in patients of advanced maternal age, recurrent pregnancy loss, and implantation failure with this technology.⁶⁸ We are hopeful that in the upcoming years, either polar body biopsy or trophectoderm biopsy will decrease the invasiveness of the procedure, and CGH or microarray technology will improve the diagnostic accuracy. Finally, even a noninvasive marker of embryonic competence may be identified to assist in selecting the ideal embryo to transfer.

Single-gene disorders can be approached by PGD with polymerase chain reaction. This technique has been used to diagnose at least 170 different conditions, cystic fibrosis and hemoglobin disorders being the most frequent indications.⁶⁹ Some other indications include many late-onset diseases such as familial polyposis coli,⁷⁰ blood group incompatibility in Rh-sensitized mothers,⁷¹ and human leukocyte antigen matching for later hematopoietic transplantation to an affected sibling.⁷²

Fertility Preservation

Advances in oncological treatment leading to better survival rates and improved public awareness of the age-related decline in fertility have led to the introduction of technical advances in oocyte cryopreservation technologies for both medical and social indications. Patients diagnosed with neoplastic or severe medical illness for which they will be treated with a gonadotoxic treatment (medical, radiological, or both) may elect to attempt to preserve their oocytes. In addition, single reproductive-age women seeking to delay their fertility may also consider oocyte cryopreservation as they attempt to fight off the age-related decline in fertility (Table 2).

Survival rates in young adults and children with malignancies have been reported to be as high as 70%.⁷³ This is in part due to the use of highly effective but gonadotoxic drugs that induce follicular atresia leading to premature ovarian failure.^{74,75} Simultaneously, the mean age of mothers in the United States has increased by 2.6 years in the last years, and it appears that education and career are important

factors in women's decisions to delay marriage and motherhood.^{76,77} Unfortunately, an important drop in fertility potential and accelerated follicular atresia are seen at the age of 37.5 years.⁷⁸

Historically, methods to preserve female fertility have included the cryopreservation of cells or tissue and surgical interventions for women undergoing pelvic radiation.⁷⁹ According to the most recent recommendations from the American Society of Reproductive Medicine and the American Society of Clinical Oncology, embryo cryopreservation is the only established option for fertility preservation in female cancer patients.^{80,81} Embryo freezing is a proven, efficient technique for fertility preservation. The downside is that on many occasions patients do not have an established relationship or neglect the use of donor sperm.⁸² Other techniques, such as oocyte and ovarian tissue cryopreservation, are currently offered to patients, but these techniques should still be considered experimental. At Reproductive Medicine Associates of New York of the Mount Sinai Medical Center, under institutional review board consent, oocyte cryopreservation may be offered to single women for attempted preservation of fertility. Especially within the context of appropriate informed consent, the preliminary results with oocyte preservation have been somewhat encouraging (Table 3).⁸³

Evaluating Ovarian Reserve

It is imperative for the physician to appropriately counsel a patient about the likelihood of achieving reproductive success prior to the onset of an IVF cycle. Prognosis is related to age, prior reproductive

It is imperative for the physician to appropriately counsel a patient about the likelihood of achieving reproductive success prior to the onset of an IVF cycle.

history, endocrine markers, and sonographic findings.

Numerous biomarkers have been studied in an attempt to find an ideal predictor of ovarian response and reproductive outcome. Basal follicular stimulating hormone (FSH) is the most frequently used hormonal assay. Optimal response is seen in patients with an FSH level below 15 IU/L,⁸⁴ and the percentage of cycles being canceled increases dramatically after 25 IU/L.⁸⁵ Even though a high FSH level is a good predictor for poor ovarian response, the patient's age should be considered before it is determined whether a patient is a candidate for IVF.⁸⁶

Table 2. Indications for Fertility Preservation.

Indication	Condition
Exposure to gonadotoxic drugs or radiation	Cancer
	Hematological disorders requiring a bone marrow transplant
Surgical menopause	Autoimmune diseases
	Prophylactic oophorectomy
	Benign ovarian disease requiring ablative surgery
Increased risk of premature ovarian failure	Turner's syndrome
	Trisomy X
	Fragile X syndrome
	Autoimmune diseases
	Environmental factors
Need or choice to delay fertility	Different social situations

NOTE: This table was adapted from Gidoni *et al.*⁷⁹

Table 3. *Fertility Preservation Techniques.*

Technique	Advantages	Disadvantages	Pregnancy Rate
Embryo cryopreservation	Established method.	Spermatozoa required.	Up to 50%
	Approved by ASRM. High success rates.	Not feasible for young single patients. Possible religious and moral objections.	
Oocyte cryopreservation	Spermatozoa not required.	ICSI required.	Variable results depending on the technique (vitrification, slow freeze, or IVM vitrification)
Ovarian tissue cryopreservation	Avoids embryo storage and disposal dilemma. Useful in young single patients. Spermatozoa not required.	Possible risks of meiotic spindle damage. Currently experimental. Surgery required.	Only 2 human births reported.
	Avoids embryo storage and disposal dilemma. Useful in young single patients. Thousands of primordial follicles preserved. Can be done on any day of the cycle. (?) Preserves endocrine function. (?) Primordial follicles are less cryosensitive.	Loss of 60% of all follicles. Graft has a limited lifespan. Experimental only.	

NOTE: This table was adapted from Gidoni *et al.*⁷⁹

Abbreviations: ASRM, American Society of Reproductive Medicine; ICSI, intracytoplasmic sperm injection; IVM, in vitro maturation.

With the understanding that FSH levels are more sensitive than luteinizing hormone (LH) to ovarian peptide feedback, our group assessed the FSH/LH ratio as a marker and indeed reported that an FSH/LH ratio greater than 3.6 is highly predictive of a poor response.⁸⁷ Basal levels of estradiol (E2) also can contribute as a predictor of ovarian response, especially when it is evaluated in non-PCOS patients with normal FSH levels. In such patients, an elevated early follicular E2 level predicts a significant increase in cancellation rates and a decrease in pregnancy rates when it is over 70 pg/mL.⁸⁸ Low inhibin B levels (<45 pg/mL) may reflect diminished ovarian reserve,⁸⁹ but the normal range is very wide, and a poor discriminatory zone has thus far limited its utility as a marker in all but the most severely compromised patients. One of the main disadvantages of the previously mentioned assays is that their secretion varies greatly during the menstrual cycle, and they have to be performed during the second or third day of the cycle. Anti-Müllerian hormone is secreted in a constant fashion by the preantral follicles during the whole menstrual cycle; this offers the advantage that it can be measured on any day of the month. This hormone correlates well with early antral follicles and therefore predicts the cohort that could

be recruited after stimulation.⁹⁰ This assay is widely used in Europe, but our experience is still limited here in the United States because it has not offered a clear advantage over the previously described tests.

Sonographic findings can also be useful in assessing ovarian reserve. The basal antral follicular count and ovarian volume measured during vaginal ultrasonography can offer accurate predictions of ovarian response.^{91,92} Some dynamic tests have been designed to further test the ability of the ovaries to respond to medication. These tests, such as the clomiphene citrate challenge test,⁹³ the gonadotropin-releasing hormone (GnRH) agonist stimulation test,⁹⁴ and the exogenous FSH ovarian reserve test,⁹⁵ are usually reserved for situations in which the previous medical history, physical examination, and other less time-consuming tests cannot determine the prognosis and the most adequate way to stimulate the patient's ovaries.

Ovarian Stimulation

The individualization of stimulation regimens, the prospective analysis of ovarian reserve, and the establishment of defined goals with respect to ovarian

response to stimulation are necessary to effectively perform COH. Many different medications and dosing

The individualization of stimulation regimens, the prospective analysis of ovarian reserve, and the establishment of defined goals with respect to ovarian response to stimulation are necessary to effectively perform COH.

schemes have been proposed over the years, and no single ovarian stimulation protocol has been shown to be the best. COH protocols need to be tailored individually, depending on the characteristics of the patient being stimulated and the type of treatment being offered.

Patients undergoing COH for IVF could be divided into 3 main groups: those that are at an increased risk of overresponding (high responders), those that are expected to have an adequate response (normal responders), and those classified as poor responders. In this context, taking into consideration the pharmacodynamics and pharmacokinetics of the medications available, we can make a decision about which stimulation protocol to use.

Young age, a high basal antral follicular count, and PCOS are all independent risk factors for developing OHSS.^{96,97} The main goal when these patients are stimulated is to recruit a reasonable number of follicles and to keep them growing in an ordered and slow fashion. Multiple studies have shown that the use of GnRH antagonists in combination with gonadotropins is a safe and effective approach in this group of patients.^{98,99} The use of oral contraceptives in the month prior to stimulation might help synchronize the oocyte cohort to be selected and promote a smoother increase in follicular size and estrogen production.¹⁰⁰ Additionally, the use of a contraceptive will be useful in scheduling these patients for their IVF cycle because they are often anovulatory and irregular.¹⁰¹

The addition of GnRH agonists has played an important role in improving ART outcomes, especially by decreasing the percentage of cancelled IVF cycles secondary to a premature luteinization.¹⁰² The use of the long protocol or down-regulation protocol, in which a GnRH agonist is started at least 1 week before stimulation, is considered the cornerstone of ovarian hyperstimulation. It is the most widely studied protocol and is commonly thought to offer the

best pregnancy and live birth rates in the normal responder.⁹⁹

It is widely accepted that a poor responder is a patient that will have a suboptimal response to stimulation. Unfortunately, no universal criteria exist to define these patients; however, at Reproductive Medicine Associates of New York, we identify a poor responder as any patient who responds to an appropriate dose of gonadotropin stimulation with less than 4 follicles or whose peak E2 level does not exceed 500 pg/mL. Patients who require an excessive amount or duration of gonadotropin stimulation may also be classified as poor responders.¹⁰⁰ If the patient has not been stimulated in the past, a poor response can be predicted with the tests reviewed in the previous section. Many protocols have been studied in an effort to optimize the quality and quantity of oocytes retrieved in this group of patients. The use of the long protocol in a patient with a diminished ovarian reserve will probably oversuppress the ovary and compromise the response to stimulation. If the GnRH agonist is started at the same time as the gonadotropins, ovarian oversuppression will be avoided, and the GnRH agonist flare effect can be used to enhance the recruitment of follicles.^{103,104} The use of an antagonist protocol will also avoid ovarian oversuppression, and it has been shown to be superior to the down-regulation protocol in poor responders.¹⁰⁰ Other adjuvants such as growth hormone, dexamethasone, dehydroepiandrosterone, and nitric oxide donors have been studied, but their use has not been shown to be advantageous.¹⁰⁵

OOCYTE RETRIEVAL AND FERTILIZATION

Oocyte Retrieval

The procedure of oocyte retrieval is standardized among clinics around the United States and the world. Under mild sedation, a vaginal ultrasound probe is used to visualize and, via a needle threaded through a guide, aspirate the contents of the follicles. Some authors have suggested using follicular flushing in order to improve oocyte retrieval, but no benefit has been seen from this more expensive and time-consuming technique.¹⁰⁶

Fertilization Methods

Although conventional insemination (the placement of approximately 50,000 sperm near each oocyte) continues to be the primary fertilization method in non-male factor IVF, ICSI is the fertilization modality

of choice in cases of severe male factor infertility. Ever since it was first performed successfully in 1992 by Palermo *et al.*,¹⁰⁷ ICSI has found widespread acceptance. ICSI can result in high fertilization and pregnancy rates, regardless of the sperm concentration, motility, or morphology, even when epididymal or testicular spermatozoa are used.^{108,109} Because of the high success rate of ICSI, some consider this technique for all cases requiring IVF. Many issues need assessment before ICSI is recommended as the standard for IVF procedures. Specific, highly technological equipment and qualified personnel are required for the procedure to be optimally performed, and previous studies in non-male factor couples have not been able to demonstrate a clear advantage of ICSI over conventional insemination.¹¹⁰ Many studies have failed to link a significant increase in the risk of adverse obstetrical outcomes, congenital malformations, and chromosomal and developmental abnormalities to ICSI in comparison with conventional insemination IVF. Further long-term studies are needed to draw definitive conclusions. For this reason, we believe that in the absence of an indication to assist fertilization, the simpler and less expensive conventional insemination procedure should be used.^{111,112}

In Vitro Maturation

In vitro maturation (IVM) is an experimental technique that offers some theoretical advantages over conventional IVF. Oocyte retrieval is performed according to a simple protocol with decreased or no hormonal stimulation. After retrieval, oocytes are matured in vitro with LH-supplemented media, and ICSI is then performed. By minimizing or eliminating ovarian stimulation, this protocol is less expensive and time-consuming; in addition, it completely avoids the risk of OHSS. IVM can be particularly useful in cancer patients undergoing fertility preservation, as it can avoid potentially toxic exposure to high doses of estradiol and may expedite the introduction of the appropriate oncological treatment. IVM success rates have been increasing over time, but despite several investigators reporting success rates as high as 21% to 54%,^{113–116} many have not achieved such success.¹¹⁷

Gamete and Embryo Cryopreservation

Sperm cryopreservation and embryo cryopreservation have been proven over the past 2 decades to be effective, safe, and established techniques. Despite having first been successful more than 20 years ago,¹¹⁸ oocyte cryopreservation is still considered experimental. The high water content of oocytes has

Despite having first been successful more than 20 years ago, oocyte cryopreservation is still considered experimental.

long presented a problem to embryologists, as oocytes have proven to be particularly susceptible to ice crystallization. Recent advances in ART, including improved culture media and optimization of cryoprotectants, however, have improved outcomes.¹¹⁹ Other problems have included a hardening of the zona pellucida following freezing, but this has been overcome by the achievement of fertilization with ICSI. Because human oocytes are viable at 37°C and no biological activity takes place at -196°C, the time of greatest danger during cryopreservation appears to be during freezing and thawing. When water becomes ice, it becomes less dense and occupies a greater volume inside the cell, increasing the pressure and damaging the intracellular organelles. Cryoprotectant substances are used to prevent ice formation and protect the cell. These substances are known to be embryotoxic, and long exposure to them might be harmful to the oocyte. Two different protocols are currently used: slow freezing and vitrification. The slow freezing method relies on low initial cryoprotectant concentrations while slowly decreasing the temperature. During vitrification, high doses of cryoprotectants are used, but the exposure is limited because freezing occurs very rapidly. Encouraging results have been reported with both techniques. Oocyte survival rates as high as 59.5% and pregnancy rates around 33% with autologous oocytes and up to 57.1% with donor oocytes have been reported with slow freezing.^{120–122} Using oocyte vitrification, Kuwayama *et al.*¹²³ reported a 90.8% survival rate with a pregnancy rate per transfer of 41.4%; using donor oocytes, Lucena *et al.*¹²⁴ reported a pregnancy rate of 56.5%.

EMBRYO SELECTION AND TRANSFER

Technological advances in IVF have led to progressive improvements in patient outcomes. One of the most critical parts of IVF has remained unchanged since the beginning, as embryologists depend primarily on the morphological appearance under inverted light microscopy to assess and select embryos for replacement. Many useful morphological classifications have been described in cleavage stage embryos, including the cleavage rate, blastomere symmetry, cytoplasmic appearance, fragmentation extent, and



Fig 2. Expanded 5-day-old blastocyst starting to hatch (courtesy of Jason Barritt, PhD).

blastomere nuclear status. In addition, extending the embryo culture from 3 to 5 days has assisted in better identifying good-prognosis embryos. Morphological grading includes assessment of the inner cell mass and trophoblast morphology and blastocoele expansion (Figure 2).¹²⁵ Although all these parameters have been correlated with implantation rates, on average, only 2 of 10 embryos implant, and this demonstrates the poor predictive value of morphological assessment alone.¹²⁶

Aside from morphological evaluation, PGS using FISH or other molecular technologies such as CGH is the only tool we have for selecting the most viable embryo.¹²⁷ Independent of the fact that pregnancy rates have not improved with PGS, routine clinical application has been impaired by the complexity, invasiveness, and high cost of the procedure. Abundant research is currently being conducted in search of less invasive and more practical techniques to complement morphological assessment for embryo selection. Most of the studies have been focusing on

Abundant research is currently being conducted in search of less invasive and more practical techniques to complement morphological assessment for embryo selection.

identifying different substances in the culture media. The presence of soluble human leukocyte antigen G in the blastocyst culture has been correlated with embryo quality and higher pregnancy rates¹²⁸; however, methodological flaws and contradictory studies that attempted to reproduce these findings have

limited its acceptance as a biomarker. Other laboratory techniques are now being applied to the field of ART, including surface-enhanced laser desorption ionization time-of-flight mass spectrometry to provide protein profiles that are specific to each individual embryo (eg, proteomic profile) to help predict their viability.¹²⁹ Metabolic activity has been the focus of several studies measuring either glucose activity^{130,131} or amino acid turnover¹³² in correlation with embryo viability and implantation capacity. Raman and near-infrared spectroscopy is now being used to profile the metabolic secretions of embryos and oocytes in culture media.¹³³ This technique, known as metabolomics (or exometabolomics), is also being correlated with embryo viability and, in the near future, may provide more practical and objective criteria for selecting the embryo or embryos to transfer.¹³⁴

Some important advances are also taking place in the field of PGD. When single-gene-defect PGD is performed, the sequence information is required prior to testing. It may be expected that with the expanded use of single nucleotide polymorphisms, linkage analysis will allow the diagnosis of any genetic disease, regardless of the availability of specific sequence information. DNA microarray technology may also provide the possibility of simultaneous testing for a causative gene, multiple linked markers, health-related genetic variability, and chromosomal abnormalities in the same biopsied blastomere.¹³⁵

IN VITRO FERTILIZATION OUTCOMES

Approximately 1 in 50 births in Sweden, 1 in 60 births in Australia, and 1 in 80 to 100 births in the United States result from IVF.¹³⁶ This demand has been met in great part by the advent of IVF-complementing techniques such as ICSI, PGD, and the use of frozen embryos and gametes. That being said, it is crucial that we remain vigilant regarding the safety and efficacy of these procedures. It has been repeatedly reported that the risk of multiple pregnancies is much higher after ART. Multiple pregnancies, either naturally or IVF-conceived, represent a greater perinatal risk. When we compare natural singleton pregnancies and IVF singleton pregnancies, the odds ratios for perinatal mortality, low birth weight, and preterm birth are 2.2, 1.7, and 1.93, respectively.¹³⁷ When the perinatal and obstetrical variables are compared between ICSI and IVF, the results are very similar.¹³⁸ With respect to birth defects, Hansen *et al.*¹³⁹ reported in 2002 that 9% of IVF babies had a birth defect versus 4.2% of non-IVF babies. The

odds ratio was 2.2 for both IVF and ICSI singletons. In 2005, the same authors performed a meta-analysis that showed a pooled odds ratio of 1.4.¹⁴⁰

Although the data support the idea that there are no recurring genomic deficits related to the fertilization of oocytes in vitro, it has recently been proposed that we should be investigating epigenetic phenomena as well. Imprinting is a rare condition sometimes linked to IVF.¹⁴¹ As an epigenetic mechanism, it can lead to alterations in DNA function without altering the DNA sequence itself. Although most genes are expressed from both parental loci simultaneously, a subset of genes are expressed only from paternal or maternal loci. The inactivated gene is the imprinted one, and this inactivation is accomplished by the process of methylation. Errors in methylation and therefore imprinting can occur during gametogenesis when germ cells undergo meiosis to become haploid.¹⁴² It is proposed that IVF could interfere with this process and therefore be associated with imprinting errors. As a result, there may be an increase in cancers due to a loss of imprinting of a growth suppressor gene. Beckwith-Wiedemann syndrome has been associated with somatic overgrowth and tumor predisposition and has been identified to have a 9-fold increase in prevalence following IVF and ICSI.¹⁴³ Angelman and Silver-Russell syndromes have also been evaluated for possible association with IVF.¹⁴⁴ The precise nature of any imprinting alteration is not clear, nor is the association with the method of conception. It is possible that the epigenetic alterations may have even been part of the etiology and not the treatment of infertility in some of these patients.

CONCLUSION

As reproductive medicine has been evolving the treatment of infertility has transitioned from medicine to surgery, and now we are experiencing a further evolution to molecular and genetic approaches. Applied technology is taking us even deeper into the structure of reproductive matter, so micromanipulation in the laboratory is the new operative theater. As the development of new tools makes assisted reproduction more efficient and safer, IVF is becoming the treatment of choice for many causes of infertility and fertility-related situations. Along the way, the goals have been the same: evidence-based and translational medicine, forward-thinking treatment, involvement of the educated patient in decision making, and, of course, the opportunity for our patients to conceive a healthy child in a safe environment for both the patient and the conceptus.

ACKNOWLEDGMENT

Lawrence Grunfeld and Tanmoy Mukherjee are acknowledged for their mentorship, guidance, and clinical expertise.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Copperman AB, DeCherney AH. Turn, turn, turn. *Fertil Steril* 2006; 85: 12–13.
2. Feinberg EC, Levens ED, DeCherney AH. Infertility surgery is dead: only the obituary remains? *Fertil Steril* 2008; 89: 232–236.
3. Centers for Disease Control. The Fertility Clinic Success Rate and Certification Act of 1992. *Fed Regist* 1998; 63: 215.
4. Andersen A, Goossens V, Gianaroli L, et al. Assisted reproductive technology in Europe, 2003. Results generated from European registries by ESHRE. *Hum Reprod* 2007; 22: 1513–1525.
5. Society for Assisted Reproductive Technology, American Society for Reproductive Medicine. Assisted reproductive technology in the United States: 2001 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology registry. *Fertil Steril* 2007; 87: 1253–1266.
6. Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo. *Lancet* 1978; 12: 366.
7. Toner JP. Progress we can be proud of: U.S. trends in assisted reproduction over the first 20 years. *Fertil Steril* 2002; 78: 943–950.
8. Habana AE, Palter SF. Is tubal embryo transfer of any value? A meta-analysis and comparison with the Society for Assisted Reproductive Technology database. *Fertil Steril* 2001; 76: 286–293.
9. Levran D, Farhi J, Nahum H, et al. Prospective evaluation of blastocyst stage transfer vs. zygote intrafallopian tube transfer in patients with repeated implantation failure. *Fertil Steril* 2002; 77: 971–977.
10. Aslan D, Elizur SE, Levron J, et al. Comparison of zygote intrafallopian tube transfer and transcervical uterine embryo transfer in patients with repeated implantation failure. *Eur J Obstet Gynecol Reprod Biol* 2005; 122: 191–194.
11. Gomel V. Reversal of tubal sterilization versus IVF on the era of assisted reproductive technology: a clinical dilemma. *Reprod Biomed Online* 2007; 15: 403–407.
12. Dubuisson JB, Bouquet de Joliniere J, Aubriot FX, et al. Terminal tuboplasties by laparoscopy: 65 consecutive cases. *Fertil Steril* 1990; 54: 401–403.
13. Filippini F, Darai E, Benifla JL, et al. Distal tubal surgery: a critical review of 104 laparoscopic distal tuboplasties. *J Gynecol Obstet Biol Reprod* 1996; 25: 471–478.
14. Dlugi AM, Reddy S, Saleh WA, et al. Pregnancy rates after operative endoscopic treatment of total

- (neosalpingostomy) or near total (salpingostomy) distal tubal occlusion. *Fertil Steril* 1994; 62: 913–920.
15. Taylor RC, Berkowitz J, McComb PF. Role of laparoscopic salpingostomy in the treatment of hydrosalpinx. *Fertil Steril* 2001; 75: 594–600.
 16. Flood JT, Grow DR. Transcervical tubal cannulation: a review. *Obstet Gynecol Surv* 1993; 48: 768–776.
 17. Valle RF. Tubal cannulation. *Obstet Gynecol Clin North Am* 1995; 22: 519–540.
 18. Patton PE, Williams TJ, Coulam CB. Microsurgical reconstruction of the proximal oviduct. *Fertil Steril* 1987; 47: 35–39.
 19. Marana R, Quagliarello J. Proximal tubal occlusion: microsurgery versus IVF—a review. *Int J Fertil* 1988; 33: 338–340.
 20. Gomel V. Microsurgery in infertility. *Int J Fertility* 1978; 23: 244–245.
 21. Mukherjee T, Copperman AB, McCaffrey C, et al. Hydrosalpinx fluid has embryotoxic effects on murine embryogenesis: a case for prophylactic salpingectomy. *Fertil Steril* 1996; 66: 851–853.
 22. Copperman AB, Wells V, Luna M, et al. Presence of hydrosalpinx correlated to endometrial inflammatory response in vivo. *Fertil Steril* 2006; 86: 972–976.
 23. Johnson NP, Mak W, Sowter MC. Surgical treatment for tubal disease in women due to undergo in vitro fertilization. *Cochrane Database Syst Rev* 2004; 3: CD002125.
 24. Strandell A, Lindhard A, Waldenstrom U, Thorburn J. Hydrosalpinx and IVF outcome: cumulative results after salpingectomy in a randomized controlled trial. *Hum Reprod* 2001; 16: 2403–2410.
 25. Cohen MA, Lindheim SR, Sauer MV. Hydrosalpinges adversely affect implantation in donor oocyte cycles. *Hum Reprod* 1999; 14: 1087–1089.
 26. Strathy JH, Molgaard CA, Coulam CB, Melton LJ. Endometriosis and infertility: a laparoscopic study of endometriosis among fertile and infertile women. *Fertil Steril* 1982; 38: 667–672.
 27. Verkauf BS. The incidence, symptoms and signs of endometriosis in fertile and infertile women. *J Fla Med Assoc* 1987; 74: 671–675.
 28. Marcoux S, Maheux R, Berhle S, for the Canadian Collaborative Group on Endometriosis. Laparoscopic surgery in infertile women with minimal or mild endometriosis. *N Engl J Med* 1998; 377: 217–222.
 29. Luciano AA, Lowney J, Jacobs SL. Endoscopic treatment of endometriosis-associated infertility: therapeutic, economic and social benefits. *J Reprod Med* 1992; 37: 573–576.
 30. Bosteels J, Van Herendael B, Weyers S, D'Hooghe T. The position of diagnostic laparoscopy in current fertility practice. *Hum Reprod Update* 2007; 13: 477–485.
 31. Tanahatog SJ, Hompes PG, Lambalk CB. Investigation of the infertile couple: should diagnostic laparoscopy be performed in the infertility work up program in patients undergoing intrauterine insemination? *Hum Reprod* 2003; 18: 8–11.
 32. Navarro J, Garrido N, Remohíacute; J, Pellicer A. How does endometriosis affect infertility? *Obstet Gynecol Clin North Am* 2003; 30: 181–192.
 33. Kennedy S, Bergqvist A, Chapron C, et al. ESHRE guideline for the diagnosis and treatment of endometriosis. *Hum Reprod* 2005; 20: 2698–2704.
 34. The Practice Committee of the American Society for Reproductive Medicine. Endometriosis and infertility. *Fertil Steril* 2006; 86: 156–160.
 35. Practice Committee of the American Society for Reproductive Medicine, Male Infertility Best Practice Committee of the American Urological Association. Report on optimal evaluation of the infertile male. *Fertil Steril* 2006; 86(suppl 1): S202–S209.
 36. Lee VM, Wong JS, Loh SK, Leong NK. Sperm motility in the semen analysis affects the outcome of superovulation intrauterine insemination in the treatment of infertile Asian couples with male factor infertility. *BJOG* 2002; 109: 115–120.
 37. Thessaloniki ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Consensus on infertility treatment related to polycystic ovary syndrome. *Fertil Steril* 2008; 89: 505–522.
 38. Tang T, Glanville J, Orsi N, et al. The use of metformin for women with PCOS undergoing IVF treatment. *Hum Reprod* 2006; 21: 1416–1425.
 39. Costello MF, Chapman M, Conway U. A systematic review and meta-analysis of randomized controlled trials on metformin co-administration during gonadotropin ovulation induction or IVF in women with polycystic ovary syndrome. *Hum Reprod* 2006; 21: 1387–1399.
 40. Haydardedeoglu B, Bagis T, Simsek E, et al. The impact of rescue in vitro fertilization converted from high-response gonadotropin intrauterine insemination cycles in terms of implantation and pregnancy rates as compared with matched controls. *Fertil Steril* 2009; 92: 137–142.
 41. Bronson RA. Sperm antibodies. *Immunol Allerg Clin North Am* 1990; 10: 165–184.
 42. Lombardo F, Gandini L, Lenzi A, Dondero F. Anti-sperm immunity in assisted reproduction. *J Reprod Immunol* 2004; 62: 101–109.
 43. Nagy ZP, Verheyen G, Liu J, et al. Results of 55 intracytoplasmic sperm injection cycles in the treatment of male-immunological infertility. *Hum Reprod* 1990; 10: 1775–1780.
 44. Dodson WC, Whitesides DB, Hughes CL Jr, et al. Superovulation with intrauterine insemination in the treatment of infertility: a possible alternative to gamete intrafallopian transfer and in vitro fertilization. *Fertil Steril* 1987; 48: 441–445.
 45. Centers for Disease Control and Prevention. 2003; assisted reproductive technology success rates. Available at: <http://www.cdc.gov/ART/ART2003>. Accessed August 2008.
 46. Pandian Z, Bhattacharya S, Vale L, Temleton A. In vitro fertilization for unexplained subfertility. *Cochrane Database Syst Rev* 2005; 18: CD003357.
 47. Karande VC, Korn A, Morris R, et al. Prospective randomized trial comparing the outcome and cost of in vitro fertilization with that of a traditional treatment algorithm as first-line therapy for couples with infertility. *Fertil Steril* 1999; 71: 468–475.
 48. Reindollar RH, Regan MM, Neumann PJ, et al. A randomized controlled trial of 503 couples assigned to conventional infertility treatment or an accelerated track to IVF: preliminary results of the Fast Track and Standard Treatment (FASTT) trial. *Fertil Steril* 2007; 88: S41.

49. Handyside AH, Kontogianni EH, Hardy K, Winston RM. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* 1990; 344: 768–770.
50. Donoso P, Staessen C, Fauser BC, Devroey P. Current value of preimplantation genetic aneuploidy screening in IVF. *Hum Reprod Update* 2007; 13: 15–25.
51. Gianaroli L, Magli C, Munne S, et al. Will preimplantation genetic diagnosis assist patients with a poor prognosis to achieve pregnancy? *Hum Reprod* 1997; 12: 1762–1767.
52. Platteau P, Staessen C, Michiels A, et al. Preimplantation genetic diagnosis for aneuploidy screening in women older than 37 years. *Fertil Steril* 2005; 84: 319–334.
53. Staessen C, Platteau P, Van Assche E, et al. Comparison of blastocyst transfer with or without preimplantation genetic diagnosis for aneuploidy screening in couples with advanced maternal age: a prospective randomized controlled trial. *Hum Reprod* 2004; 19: 2849–2858.
54. Munné S, Magli C, Cohen J, et al. Positive outcome after preimplantation diagnosis of aneuploidy in human embryos. *Hum Reprod* 1999; 14: 2191–2199.
55. Munné S, Sandalinas M, Escudero T, et al. Improved implantation after preimplantation genetic diagnosis of aneuploidy. *Reprod Biomed Online* 2003; 7: 91–97.
56. Twisk M, Mastenbroek S, van Wely M, et al. Preimplantation genetic screening for abnormal number of chromosomes (aneuploidies) in in vitro fertilization or intracytoplasmic sperm injection. *Cochrane Database Syst Rev* 2006; 2: CD005291.
57. Mastenbroek S, Twisk M, van Echten-Arends J, et al. In vitro fertilization with preimplantation genetic screening. *N Engl J Med* 2007; 357: 9–17.
58. Menasha J, Levy B, Hirschhorn K, Kardon NB. Incidence and spectrum of chromosome abnormalities in spontaneous abortions: new insights from a 12-year study. *Genet Med* 2005; 7: 251–263.
59. Pellicer A, Rubio C, Vidal F, et al. In vitro fertilization plus preimplantation genetic diagnosis in patients with recurrent miscarriage: an analysis of chromosome abnormalities in human preimplantation embryos. *Fertil Steril* 1999; 71: 1033–1039.
60. Platteau P, Staessen C, Michiels A, et al. Preimplantation genetic diagnosis for aneuploidy screening in patients with unexplained recurrent miscarriages. *Fertil Steril* 2005; 83: 393–397.
61. Munné S, Chen S, Fischer J, et al. Preimplantation genetic diagnosis reduces pregnancy loss in women aged 35 years and older with a history of recurrent miscarriages. *Fertil Steril* 2005; 84: 331–335.
62. Rubio C, Rodrigo L, Perez-Cano I, et al. FISH screening of aneuploidies in preimplantation embryos to improve IVF outcome. *Reprod Biomed Online* 2005; 11: 497–506.
63. Kahraman S, Bahce M, Samir H, et al. Healthy births and ongoing pregnancies obtained by preimplantation genetic diagnosis in patients with advanced maternal age and recurrent implantation failure. *Hum Reprod* 2000; 15: 2003–2007.
64. Munné S, Sandalinas M, Escudero T, et al. Improved implantation after preimplantation genetic diagnosis of aneuploidy. *Reprod Biomed Online* 2003; 7: 91–97.
65. Calogero A, Burello N, De Palma A, et al. Sperm aneuploidy in infertile men. *Reprod Biomed Online* 2003; 6: 310–317.
66. Gianaroli L, Magli MC, Cavallini G, et al. Frequency of aneuploidy in sperm from patients with extremely severe male factor infertility. *Hum Reprod* 2005; 20: 2140–2152.
67. Verlinsky Y, Tur-Kaspa I, Cieslak J, et al. Preimplantation testing for chromosomal disorders improves reproductive outcome of poor-prognosis patients. *Reprod Biomed Online* 2005; 11: 219–225.
68. Practice Committee of the Society for Assisted Reproductive Technology, Practice Committee of the American Society for Reproductive Medicine. Preimplantation genetic testing: a practice committee opinion. *Fertil Steril* 2007; 88: 1497–1504.
69. Kuliev A, Verlinsky Y. Symposium: genetic and epigenetic aspects of assisted reproduction. Preimplantation genetic diagnosis: technological advances to improve accuracy and range. *Reprod Biomed Online* 2008; 16: 532–538.
70. Ao A, Wells D, Handyside AH, et al. Preimplantation genetic diagnosis of inherited cancer: familial adenomatous polyposis coli. *J Assist Reprod Genet* 1998; 15: 140–144.
71. Seeho SKM, Burton G, Leigh D, et al. The role of preimplantation genetic diagnosis in the management of severe rhesus alloimmunization: first unaffected pregnancy: case report. *Hum Reprod* 2005; 20: 697–701.
72. Verlinsky Y, Rechitsky S, Schoolcraft W, et al. Preimplantation diagnosis for Fanconi anemia combined with HLA matching. *JAMA* 2001; 285: 3130–3133.
73. Robison LL. Methodologic issues in the study of second malignant neoplasms and pregnancy outcomes. *Med Pediatr Oncol* 1996; 1: 41–44.
74. Rivkees SA, Crawford JD. The relationship of gonadal activity and chemotherapy-induced gonadal damage. *JAMA* 1988; 259: 2123–2125.
75. Lobo RA. Potential options for preservation of fertility in women. *N Engl J Med* 2005; 353: 64–73.
76. Chen R, Morgan SP. Recent trends in the timing of first births in the United States. *Demography* 1991; 28: 513–533.
77. Caucutt E, Knowles J, Guner N, et al. Why do women wait? Matching, wage inequality, and the incentives for fertility delay. *Rev Econom Dynam* 2002; 5: 815–855.
78. Faddy MJ, Gosden RG, Gougeon A, et al. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. *Hum Reprod* 1992; 7: 1342–1346.
79. Gidoni Y, Holzer H, Tulandi T, Tan SL. Fertility preservation in patients with non-oncological conditions. *Reprod Biomed Online* 2008; 16: 792–800.
80. American Society for Reproductive Medicine. Ovarian tissue and oocyte cryopreservation. *Fertil Steril* 2006; 86: 142–147.
81. Lee SJ, Schover LR, Partridge AH, et al. American Society of Clinical Oncology recommendations on fertility preservation in cancer patients. *J Clin Oncol* 2006; 24: 2917–2931.
82. Deepinder F, Agarwal A. Technical and ethical challenges of fertility preservation in young cancer patients. *Reprod Biomed Online* 2008; 16: 784–791.

83. Barritt J, Luna M, Duke M, et al. Report of four donor-recipient oocyte cryopreservation cycles resulting in high pregnancy and implantation rates. *Fertil Steril* 2007; 87: 189.e13–189.e17.
84. Scott RT, Toner JP, Muasher SJ, et al. Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. *Fertil Steril* 1989; 51: 651–654.
85. Toner JP, Philput CB, Jones GS, Muasher SJ. Basal follicle-stimulating hormone level is a better predictor of in vitro fertilization performance than age. *Fertil Steril* 1991; 55: 784–791.
86. Luna M, Grunfeld L, Mukherjee T, et al. Moderately elevated levels of basal follicle-stimulating hormone in young patients predict low ovarian response, but should not be used to disqualify patients from attempting in vitro fertilization. *Fertil Steril* 2007; 87: 782–787.
87. Mukherjee T, Copperman AB, Lapinski R, et al. An elevated day three follicle-stimulating hormone:luteinizing hormone ratio (FSH:LH) in the presence of a normal day 3 FSH predicts a poor response to controlled ovarian hyperstimulation. *Fertil Steril* 1996; 65: 588–593.
88. Smotrich DB, Widra EA, Gindoff PR, et al. Prognostic value of day 3 estradiol on in vitro fertilization outcome. *Fertil Steril* 1995; 64: 1136–1140.
89. Seifer DB, Lambert-Messerlian G, Hogan JW, et al. Day 3 serum inhibin-B is predictive of assisted reproductive technologies outcome. *Fertil Steril* 1997; 67: 110–114.
90. Fanchin R, Schonäuer LM, Righini C, et al. Serum anti-Müllerian hormone is more strongly related to ovarian follicular status than serum inhibin B, estradiol, FSH and LH on day 3. *Hum Reprod* 2003; 18: 323–327.
91. Chang MY, Chiang CH, Hsieh TT, et al. Use of the antral follicle count to predict the outcome of assisted reproductive technologies. *Fertil Steril* 1998; 69: 505–510.
92. Syrop CH, Willhoite A, Van Voorhis BJ. Ovarian volume: a novel outcome predictor for assisted reproduction. *Fertil Steril* 1995; 64: 1167–1171.
93. Navot D, Rosenwaks Z, Margalioth EJ. Prognostic assessment of female fecundity. *Lancet* 1987; 2: 645–647.
94. Winslow KL, Toner JP, Brzyski RG, et al. The gonadotropin-releasing hormone agonist stimulation test—a sensitive predictor of performance in the flare-up in vitro fertilization cycle. *Fertil Steril* 1991; 56: 711–717.
95. Fanchin R, de Ziegler D, Olivennes F, et al. Exogenous follicle stimulating hormone ovarian reserve test (EFORE): a simple and reliable screening test for detecting “poor responders” in in-vitro fertilization. *Hum Reprod* 1994; 9: 1607–1611.
96. MacDougall MJ, Tan SL, Balen A, Jacobs HS. A controlled study comparing patients with and without polycystic ovaries undergoing in-vitro fertilization. *Hum Reprod* 1993; 8: 233–237.
97. Enskog A, Henriksson M, Unander M, et al. Prospective study of the clinical and laboratory parameters of patients in whom ovarian hyperstimulation syndrome developed during controlled ovarian hyperstimulation for in vitro fertilization. *Fertil Steril* 1999; 71: 808–814.
98. Ragni G, Vegetti W, Riccaboni A, et al. Comparison of GnRH agonists and antagonists in assisted reproduction cycles of patients at high risk of ovarian hyperstimulation syndrome. *Hum Reprod* 2005; 20: 2421–2425.
99. Al-Inany HG, Abou-Setta AM, Aboulghar M. Gonadotropin-releasing hormone antagonists for assisted conception. *Cochrane Database Syst Rev* 2006; 3: CD001750.
100. Copperman AB. Antagonists in poor-responder patients. *Fertil Steril* 2003; 80: S16–S24.
101. Meldrum D, Scott R, Levy MJ, et al. A pilot study to assess oral contraceptive (OC) pretreatment in women undergoing controlled ovarian hyperstimulation (COH) in ganirelix acetate cycles. *Fertil Steril* 2002; 78O:S176.
102. Surrey ER, Schoolcraft WB. Evaluating strategies for improving ovarian response of the poor responder undergoing assisted reproductive techniques. *Fertil Steril* 2000; 73: 667–676.
103. Howles CM, Macnamee MC, Edwards RG. Short term use of an LHRH agonist to treat poor responders entering an IVF programme. *Hum Reprod* 1987; 2: 655–656.
104. Marcus SF, Brisden PR, Macnamee M, et al. Comparative trial between an ultra short and long protocol of LHRH agonist for ovarian stimulation in IVF. *Hum Reprod* 1993; 8: 238–243.
105. Tarlatzis BC, Zepiridis L, Grimbizis G, Bontis J. Clinical management of low ovarian response to stimulation for IVF: a systematic review. *Hum Reprod Update* 2003; 9: 61–76.
106. Levens ED, Whitcomb BW, Payson MD, Larsen FW. Ovarian follicular flushing among low-responding patients undergoing assisted reproductive technology. *Fertil Steril* 2009; 91(suppl): 1381–1384.
107. Palermo G, Joris H, Devroey P, Van Steirteghem AC. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. *Lancet* 1992; 340: 17–18.
108. Devroey P, Liu J, Nagy Z, et al. Normal fertilization of human oocytes after testicular sperm extraction and intracytoplasmic sperm injection. *Fertil Steril* 1994; 62: 639–641.
109. Nagy ZP, Liu J, Jassenswillen C, et al. Using ejaculated, fresh, and frozen-thawed epididymal and testicular spermatozoa gives rise to comparable results after intracytoplasmic sperm injection. *Fertil Steril* 1995; 63: 808–815.
110. Fishel S, Aslam I, Lisi F, et al. Should ICSI be the treatment of choice for all cases of in-vitro conception? *Hum Reprod* 2000; 15: 1278–1283.
111. Baker HWG, Liu DY, Bourne H, et al. Diagnosis of sperm defects in selecting patients for assisted fertilization. *Hum Reprod* 1993; 8: 1779–1780.
112. Tucker M, Wiker S, Massey J. Rational approach to assisted fertilization. *Hum Reprod* 1993; 8: 1778.
113. Mikkelsen AL, Smith SD, Lindenberg S. In-vitro maturation of human oocytes from regularly menstruating women may be successful without follicle stimulating hormone priming. *Hum Reprod* 1999; 14: 1847–1851.
114. Chian RC, Buckett WM, Tan SL. In-vitro maturation of human oocytes. *Reprod Biomed Online* 2004; 8: 148–166.

115. Lin YH, Hwang JH, Huang LW, et al. Combination of FSH priming and hCG priming for in-vitro maturation of human oocytes. *Hum Reprod* 2003; 18: 1632–1636.
116. Söderström-Anttila V, Makinen S, Tuuri T, et al. Favorable pregnancy results with insemination of in vitro matured oocytes from unstimulated patients. *Hum Reprod* 2005; 20: 1534–1540.
117. Suikkari AM, Söderström-Anttila V. In-vitro maturation of eggs: is it really useful? *Best Pract Res Clin Obstet Gynaecol* 2007; 21: 145–155.
118. Chen C. Pregnancy after human oocyte cryopreservation. *Lancet* 1986; 2: 884–886.
119. Jain JK, Paulson RJ. Oocyte cryopreservation. *Fertil Steril* 2006; 83: 1037–1046.
120. Fabbri R, Porcu E, Marsella T, et al. Human oocyte cryopreservation: new perspectives regarding oocyte survival. *Hum Reprod* 2001; 16: 411–416.
121. Chen SU, Lien YR, Chen HF, et al. Observational clinical follow-up of oocyte cryopreservation using a slow-freezing method with 1,2-propanediol plus sucrose followed by ICSI. *Hum Reprod* 2005; 20: 1975–1980.
122. Fosas N, Marina F, Torres PJ, et al. The births of five Spanish babies from cryopreserved donated oocytes. *Hum Reprod* 2003; 18: 1417–1421.
123. Kuwayama M, Vajita G, Kato O, Leibo SP. Highly efficient vitrification method for cryopreservation of human oocytes. *Reprod Biomed Online* 2005; 11: 300–308.
124. Lucena E, Bernal DP, Lucena C, et al. Successful ongoing pregnancies after vitrification of oocytes. *Fertil Steril* 2006; 85: 108–111.
125. Rienzi L, Ubaldi F, Iacobelli M, et al. Significance of morphological attributes of the early embryo. *Reprod Biomed Online* 2005; 10: 669–681.
126. Nagy ZP. Symposium: innovative techniques in human embryo viability assessment. *Reprod Biomed Online* 2008; 17: 451–453.
127. Munné S, Chen S, Colls P, et al. Maternal age, morphology, development and chromosome abnormalities in over 6000 cleavage-stage embryos. *Reprod Biomed Online* 2007; 14: 628–634.
128. Warner CM, Lampton PW, Newmark JA, et al. Soluble human leukocyte antigen-G and pregnancy success. *Reprod Biomed Online* 2008; 17: 470–485.
129. Katz-Jaffe MG, Gardner DK. Can proteomics help to shape the future of human assisted conception? *Reprod Biomed Online* 2008; 17: 497–501.
130. Gardner DK, Leese HJ. Assessment of embryo viability prior to transfer by the noninvasive measurement of glucose uptake. *J Exp Zool* 1987; 242: 103–105.
131. Hardy K, Hooper MA, Handyside AH, et al. Non-invasive measurement of glucose and pyruvate uptake by individual human oocytes and preimplantation embryos. *Hum Reprod* 1989; 4: 188–191.
132. Sturmey RG, Brison DR, Leese HJ. Assessing embryo viability by measurement of amino acid turnover. *Reprod Biomed Online* 2008; 17: 486–496.
133. Scott R, Seli E, Miller K, et al. Noninvasive metabolomic profiling of human embryo culture media using Raman spectroscopy predicts embryonic reproductive potential: a prospective blinded pilot study. *Fertil Steril* 2008; 90: 77–83.
134. Nagy ZP, Sakkas D, Behr B. Non-invasive assessment of embryo viability by metabolomic profiling of culture media ('metabolomics'). *Reprod Biomed Online* 2008; 17: 502–507.
135. Kuliev A, Verlinsky Y. Preimplantation genetic diagnosis: technological advances to improve accuracy and range of applications. *Reprod Biomed Online* 2008; 16: 532–538.
136. Van Voorhis BJ. In vitro fertilization. *N Engl J Med* 2007; 356: 379–386.
137. Bower C, Hansen M. Assisted reproductive technologies and birth outcomes: overview of recent systematic reviews. *Reprod Fertil Dev* 2005; 17: 329–333.
138. Ombelet W, Cadron I, Gerris J, et al. Obstetric and perinatal outcome of 1655 ICSI and 3974 IVF singleton and 1102 ICSI and 2901 IVF twin births: a comparative analysis. *Reprod Biomed Online* 2005; 11: 76–85.
139. Hansen M, Kurinczuk JJ, Bower C, et al. The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *N Engl J Med* 2002; 346: 725–730.
140. Hansen M, Bower C, Milne E, et al. Assisted reproductive technologies and the risk of birth defects: a systematic review. *Hum Reprod* 2005; 20: 328–338.
141. Paoloni-Giacobino A. Implications of reproductive technologies for birth and developmental outcomes: imprinting defects and beyond. *Expert Rev Mol Med* 2006; 8: 1–14.
142. Luedi PP, Hartemink AJ, Jirtle RL. Genome-wide prediction of imprinted murine genes. *Genome Res* 2005; 15: 875–884.
143. Halliday J, Oke K, Breheny S, et al. Beckwith-Wiedemann syndrome and IVF: a case control study. *Am J Hum Genet* 2004; 75: 526–528.
144. Sutcliffe AG, Peters CJ, Bowdin S, et al. Assisted reproductive therapies and imprinting disorders: a preliminary British survey. *Hum Reprod* 2006; 21: 1009–1011.

Noninvasive Prenatal Diagnosis: Past, Present, and Future

Christian Litton, MD, Joanne Stone, MD, Keith Eddleman, MD, and
Men-Jean Lee, MD

Mount Sinai School of Medicine, New York, NY

OUTLINE

NONINVASIVE PRENATAL DIAGNOSIS: OVERVIEW

Whole Fetal Cells
Cell-Free Fetal DNA and
Fetal Messenger RNA

APPLICATIONS OF NONINVASIVE PRENATAL DIAGNOSIS

Fetal Aneuploidies
Fetal Sex Determination
Determination of Rh Status
Antenatal Diagnosis and Management
of High-Risk Pregnancies

CONCLUSION

markers (fetal cells, cell-free fetal DNA, and fetal messenger RNA) for diagnostic and screening purposes is now being developed. The scope of noninvasive prenatal diagnosis is not limited to only the diagnosis of fetal genetic traits and aneuploidies. Recently, researchers have focused their investigations on the role of cell-free fetal DNA and fetal messenger RNA in preeclampsia, intrauterine growth restriction, and preterm labor. These biomarkers, the result of inherent placental dysfunction or the byproducts of placental trophoblastic apoptosis, may allow for improvements in the diagnosis and management of high-risk pregnancies. *Mt Sinai J Med* 76:521–528, 2009. © 2009 Mount Sinai School of Medicine

Key Words: cell-free fetal DNA, fetal cells, fetal messenger RNA, noninvasive prenatal diagnosis.

ABSTRACT

The presence of fetal cells in the maternal circulation was first noted by Georg Schmorl when he documented the presence of multinucleated syncytial giant cells of placental origin in the lung tissue of women who had died from complications of eclampsia. In the intervening century, advances in cellular and molecular biology further elucidated both the physiology and pathophysiology of communication within the fetomaternal unit. This concept is at the foundation of the rapidly expanding field of noninvasive prenatal diagnosis. However, the clinical utility of this phenomenon had been limited until the presence of cell-free fetal DNA circulating in the maternal plasma was reported in 1997 and fetal messenger RNA was demonstrated to circulate in the maternal plasma in 2000. These circulating nucleic acids are found free-floating in the maternal plasma, unencumbered by a surrounding fetal cell. The analysis of these 3 fetal

The presence of fetal cells in the maternal circulation was first appreciated by Georg Schmorl in his seminal article published in Germany in 1893.¹ Schmorl documented the presence of multinucleated syncytial giant cells of placental origin in the lung tissue of 17 women who had died from complications of eclampsia. Schmorl recognized not only the presence of this communication between the placental circulation and the maternal circulation, but also its amplification in gestations affected by eclampsia. In the intervening century, advances in cellular and molecular biology further elucidated both the physiology and pathophysiology of communication within the fetomaternal unit. This concept is at the foundation of the rapidly expanding field of noninvasive prenatal diagnosis (NIPD).

Even though the presence of fetal cells in the maternal circulation has been documented for over a century, the clinical application of this phenomenon has been limited until quite recently. The field of NIPD has progressed rapidly since Lo and colleagues² observed the presence of cell-free fetal DNA (cffDNA) circulating in the maternal plasma in 1997. Subsequently, fetal messenger RNA (fmrRNA) was demonstrated to circulate in the maternal plasma

Address Correspondence to:

Men-Jean Lee
Mount Sinai School of Medicine
New York, NY
Email: menjean.lee@mssm.edu

by Poon and colleagues³ in 2000. These circulating nucleic acids are found free-floating in the maternal plasma, unencumbered by a surrounding fetal cell. The analysis of these 3 fetal markers (fetal cells, cffDNA, and fmRNA) for diagnostic and screening purposes has only recently been realized in the clinical setting.

The scope of NIPD is not limited to only the diagnosis of fetal genetic traits and aneuploidies. Recently, researchers have focused their investigations on the roles of cffDNA and fmRNA in preeclampsia (PEC), fetal growth restriction (FGR), and preterm labor (PTL). These biomarkers, the result of inherent placental dysfunction or the byproducts of placental trophoblastic apoptosis, may allow for improvements in the diagnosis and management of high-risk pregnancies.

NONINVASIVE PRENATAL DIAGNOSIS: OVERVIEW

The utility of NIPD depends on our ability to separate and analyze fetal genetic material that has entered the maternal circulation. Two major approaches have emerged in the analysis of fetal genetic materials.

Whole Fetal Cells

The first is the isolation of whole fetal cells that have entered the maternal circulation. Whole fetal cells, when isolated, offer a source of the entire fetal genome. This source is free of any confounding maternal genetic material. This is of particular importance when we are analyzing inherited genetic traits that are passed on via Mendelian inheritance as we must distinguish a mutant allele of maternal cell origin from the fetal allele of maternal origin.⁴

The main obstruction to the analysis of fetal cells remains the ability to efficiently and effectively isolate them from the maternal circulation. Fetal cells occur on the order of approximately 1 cell per milliliter of whole maternal blood.⁴ In 2002, the National Institute of Child Health and Human Development Fetal Cell Isolation Study (NIFTY) sought to address this problem by examining automated cell separation methods, including magnetic cell sorting and fluorescent activated cell sorting. However, the sensitivity of detection was approximately 50% on average, which is far below an acceptable or useful level.⁵

In addition to the impediments in isolating whole fetal cells, the analysis of these cells is complicated by a high level of allele dropout. In

Mendelian inheritance, allele dropout refers to the nonamplification of 1 of the alleles (either maternal or paternal). Recent studies have demonstrated rates of allele dropout as high as 50%. This has the potential to produce unacceptably high rates of false-negative results when we are looking for a specific genetic sequence. Such a high rate of allele dropout would become problematic if this test were to be used in the clinical arena. This would require the examination of at least 5 to 6 fetal cells in order to offset the dropout rate.⁴

Another impediment to the use of whole fetal cells is cellular apoptosis upon their entrance into the maternal circulation. It has been hypothesized that as fetal erythroblasts (the most commonly isolated whole fetal cells) transition into the oxygen-rich maternal circulation, the process of apoptosis begins. This too makes these cells particularly resistant to molecular evaluation.

Despite these impediments, research using whole fetal cells continues today. Experts agree that future research must focus on efficient and automated methods of isolation of fetal cells in the maternal circulation. Only after this obstacle is overcome may the focus be shifted to the extraction of relevant genetic material.

Cell-Free Fetal DNA and Fetal Messenger RNA

cffDNA and fmRNA are short segments of fetal nucleic acids found circulating in the maternal plasma and serum. These segments are observed in their

*cffDNA and fmRNA are short
segments of fetal nucleic acids
found circulating in the maternal
plasma and serum.*

unbound form, free from their usual cellular constituents. The presence of cffDNA in the maternal circulation was not demonstrated until 1997.² Three years later, fmRNA was demonstrated in the maternal plasma.³ These 2 discoveries have been instrumental in advancing the current state of NIPD. It is important to note that the origin of cffDNA and fmRNA has been shown to be the placental trophoblasts and not the actual fetal unit, as was initially hypothesized. This fact becomes important when we are discussing the implications of molecular testing for pregnancy and is discussed later in this review.

In its entirety, cffDNA accounts for 3% to 5% of the total circulating free DNA in the maternal

plasma. This fact makes the determination of specific fetal loci difficult. To quantify specific fetal loci, 2 methods are typically employed: identification of short tandem repeats and single nucleotide polymorphisms. Short tandem repeats and single nucleotide polymorphisms are nucleic acid sequences that can be identified as unique to the fetal DNA under investigation. However, because of the scarcity of cfDNA, these techniques are quite cumbersome and inefficient.⁴

This obstacle has nearly been overcome by the use of fmRNA. By the use of fmRNA it is possible to isolate "placenta-specific mRNA species not expressed by any maternal tissue."⁴ In doing so, researchers are able to focus solely on the fetal specific gene in question while ignoring the "maternal background noise."⁴ Currently, trisomy 21 has been detected with a sensitivity of 90% and a specificity of 96% via fmRNA in small studies.⁶ In clinical practice, fetal sex and RhD determination currently use fmRNA.

APPLICATIONS OF NONINVASIVE PRENATAL DIAGNOSIS

Fetal Aneuploidies

The holy grail of NIPD remains the detection of fetal aneuploidies.⁷ Most research to this day aimed at answering this clinical issue has been carried out by those working on the retrieval and analysis of whole fetal cells. Whole fetal cells retrieved from the maternal circulation contain the entire fetal genome and are therefore invaluable in analyzing for the presence of aneuploidies. In the analysis of cfDNA and fmRNA, it was previously thought that the presence of extraneous or inadequate genetic material would hinder the accurate diagnosis of fetal aneuploidies. This does not seem to be the case as the more current literature supports the use of cfDNA and fmRNA in the detection of aneuploidy.⁴

NIFTY, published in 2002, focused solely on whole fetal cells. It established that the current enrichment and analysis technologies of circulating fetal cells were not operational on a clinical level. In addition, the sensitivity of aneuploidy detection was comparable to that of single-marker serum screening, far from the gold standards of chorionic villus sampling and amniocentesis.⁵

In a follow-up to NIFTY, the European Union-funded Special Non-Invasive Advances in Fetal and Neonatal Evaluation (SAFE) Network attempted to address the current limitations in the

enrichment of fetal cells. The study was aimed to optimize fetal cell enrichment by relying heavily on the automated separation of fetal cells from the maternal circulation. Fetal cell separation was optimized with either magnetic or fluorescence cell sorters. This study achieved far greater success in isolating whole fetal cells than previous studies.⁸

The next hurdle was noted to be in the amplification and analysis of the isolated fetal cells. Isolated fetal cells (typically fetal erythroblasts) are particularly refractory to the current methods of fluorescence in situ hybridization analysis. In addition, researchers also encountered an excessively high level of allele dropout when performing single-cell polymerase chain reaction (PCR). In conjunction, these obstacles impede the use of whole fetal cells in the diagnosis of aneuploidy at this time.

More current research has concentrated on the use of fmRNA in the determination of aneuploidy. Lo and colleagues² recently published their research on the use of the placenta-specific 4 (PLAC4) gene in identifying trisomy 21. The PLAC4 gene is known to be expressed only in placental tissue. It is also located on chromosome 21 in the human genome. In the study, the isolated fmRNA was amplified and analyzed. The ratio of the alleles present was then enumerated. If the gene were present in its heterozygous form, the ratio of the alleles would be 1:1. If the ratio were noted to be 2:1, it could be assumed that there was an extra copy of 1 of the alleles and hence that the fetus had trisomy 21. The main disadvantage of this technique was that the gene had to be present in its heterozygous form. If the sample were taken from a homozygous specimen, only a single allele would be noted during the analysis.

In order to overcome this restriction of zygosity, it became necessary to implement novel detection methods. In 2007, Fan and Quake⁹ also demonstrated the utility of digital polymerase chain reaction (dPCR) in detecting aneuploidy. Briefly, dPCR is a technique by which the amount of DNA in a sample can be quantified via the counting of the amplifications of single molecules. It is, therefore, possible to accurately enumerate minute changes in template copy numbers. The procedure of dPCR, as described in the experimental design, was noted to be both simple and rapid. The entire diagnostic procedure can be completed within a few hours. This is considerably quicker than the conventional karyotype analysis. Additionally, Fan and Quake showed that it is possible with dPCR to detect a fetal aneuploidy in the presence of maternal contamination or fetal chromosomal mosaicism. However, the main

Fan and Quake showed that it is possible with dPCR to detect a fetal aneuploidy in the presence of maternal contamination or fetal chromosomal mosaicism.

limitation is that it requires a large starting quantity of DNA material equivalent to what is found in typical amniocentesis/chorionic villus sampling. Ideally, scaling down this methodology to a specimen amount containing fetal DNA that is equivalent to what is available through noninvasive techniques would be make this technology more clinically applicable.⁹

In a follow-up study, Lo *et al.*¹⁰ described a similar procedure. However, the starting genetic material used in the experimental design was cffDNA. Using dPCR, they demonstrated that it was no longer necessary for the gene to be present in its heterozygous form in order for the trisomy to be identified. In addition, the use of small quantities of cffDNA demonstrated the feasibility of using a noninvasive sample acquisition method. Although the small sample size in the study was small, the sensitivity of detection was noted to be 100%.

Recent studies by Fan *et al.*¹¹ have made use of high-throughput shotgun sequencing of DNA. In shotgun sequencing, a DNA sample is randomly fractionated into smaller segments. These small segments are then sequenced with chain terminations. Several fraction/sequence rounds are performed to obtain multiple DNA reads. These DNA reads are then assembled into a continuous DNA segment with the aid of complicated informatics computer technology. This technique allows for more efficient sequencing of small quantities of DNA. They demonstrated that an accurate diagnosis of fetal aneuploidy with shotgun sequencing of cffDNA can be made. This method allows for universal polymorphism-independent assays for noninvasive testing with small amounts of cffDNA from maternal plasma. In addition, as the quantity of cffDNA in the maternal circulation composes approximately 3% of the total free DNA, Fan *et al.* described a calculation

Shotgun sequencing of DNA allows for universal polymorphism-independent assays for noninvasive testing with small amounts of cffDNA from maternal plasma.

to assess the number of array compartments required to assume a predetermined confidence interval. This allows for an exact calculation for the sensitivity of the test, which can be increased by mere changes in the scale of the assay. This is done in order to overcome the relatively small quantity of cffDNA in the maternal circulation.¹¹

Another consideration when one is working with cffDNA is the difference in the lengths of the DNA segments to be analyzed. Lun *et al.*¹² described an experimental process known as digital nucleic acid size selection, by which nucleic acids of various lengths can be separated. This technique is particularly useful in the enrichment of cffDNA from the background maternal DNA. This technique exploits the fact that cffDNA segments are typically shorter in length than circulating maternal DNA. It is, therefore, possible to enrich the percentage of cffDNA from the background pool, and this can lead to more robust measurements of the target genetic material. This technique, combined with dPCR, has allowed researchers to push forward in their efforts to attain NIPD of a variety of disease states.

Despite the recent progress made in the area of NIPD, more work is required prior to its application in the clinical setting. The determination of fetal aneuploidies remains, at this time, a nonclinical experimental protocol. Small studies have shown that fetal aneuploidies can be determined with great confidence by NIPD. Given its implications, we need the results from large, ongoing multicenter trials before the introduction of NIPD for the determination of aneuploidy for clinical use.

Fetal Sex Determination

The diagnosis of fetal sex was one of the earliest clinically used tests in the arena of NIPD, particularly for the prenatal diagnosis of X-linked inherited disorders. Currently, an accurate and realizable diagnosis of fetal sex can be made as early as 7 weeks' gestation. The testing involves the identification of cffDNA circulating in maternal plasma. The clinical indications for determining fetal sex include X-linked genetic disorders, metabolic conditions associated with genital ambiguity, and confirmation of fetal sex. At this time, certain European countries are offering NIPD for sex determination as a clinical service.¹³

The appropriate and successful treatment of disease states such as fetal congenital adrenal hyperplasia relies on the early determination of gender. The treatment of a fetus with congenital adrenal hyperplasia must begin as early as 9 weeks'

gestation to prevent the virilization of a female fetus. Typically, treatment with high-dose glucocorticoids is begun as soon as pregnancy is diagnosed and is discontinued once the fetus is found to be male by chorionic villus sampling, which cannot easily be performed prior to 10 weeks' gestation. With earlier sex determination, the unnecessary steroid treatment of over half of these patients would be avoided.

The basis of current fetal sex determination lies in the fact that in a female fetus, the Y chromosome should not be detected in the maternal plasma.

The basis of current fetal sex determination lies in the fact that in a female fetus, the Y chromosome should not be detected in the maternal plasma.

In the absence of any detectable Y chromosome, the diagnosis of a female fetus is presumed. Therefore, it is important that the absence of a detectable Y chromosome in the serum be confirmed to represent a female fetus and not simply an error in the testing itself. This may be performed with ultrasound in selected cases. The accuracy of fetal sex determination by the PCR analysis of cfDNA ranges from 87% to 100% in the literature.⁸ Given this inaccuracy, the testing is still considered experimental in the United States. The tradeoff in clinical practice lies in the ability to determine fetal sex at a much earlier gestational age than is feasible with the currently available technology.

There are 2 goals of future research in the field of fetal sex determination. The first is focused on attaining a sex-independent fetal marker. Therefore, the diagnosis of fetal sex would then lie not in the absence of a Y chromosome but rather in the presence of a specific genomic marker. The second is to improve the diagnostic accuracy of the current technique. With the introduction of novel molecular techniques, the accuracy of diagnosis has vastly improved. However, further advances are required before this technology can be made available for general clinical use.

Determination of Rh Status

RhD incompatibility between a mother and a fetus may lead to maternal alloimmunization and subsequent pregnancy complications. Hemolytic disease of the newborn can occur in an RhD-positive fetus if the mother has previously been alloimmunized. The current recommendations call

for immunoprophylaxis with RhoGAM (anti-RhD antibody) for all RhD-negative mothers. However, if the fetal blood type is also RhD-negative, there is no risk of anti-D alloimmunization. Consequently, a large portion of the population is subjected to the risks of RhoGAM, including bloodborne disease transmission and allergic reactions. The financial burden on the health care system of unnecessarily treating so many mothers must also be considered.¹¹

Until recently, the only method to determine the fetal RhD status involved invasive testing, which potentially can sensitize the pregnant woman further. With the introduction of NIPD, it is now possible to determine the fetal RhD status and significantly change the manner in which these pregnancies are managed. The methodology for detecting the presence of an Rh-positive fetus is similar to that for detecting the fetal Y chromosome. In fetal sex determination, the Y chromosome is sought in the maternal XX background. Similarly, in fetal RhD determination, the presence of the fetal RhD gene is identified in the background of an RhD-negative maternal genome. The majority of studies thus far have used PCR in the detection of cfDNA because of its ease of use and accuracy.

In the population, there exists genetic variation in the expression of the RhD gene. These differences are particularly apparent when different ethnicities are being compared. For instance, the *RHD-CE-D* gene, present in 22% of D-negative African Americans, can lead to false-positive genotyping results when PCR is being used. In addition, some patients carry what is known as an RhD pseudogene. Patients with the RhD pseudogene have missense mutations at specific exons.⁹ When these patients are being evaluated, they typically are defined as RhD-positive by standard blood bank testing. However, the maternal genome is RhD-negative, and consequently, these patients may develop anti-D antibodies. In order to address this issue, the SAFE Network has recommended the assessment of particular exons when the RhD gene is being evaluated. Using the recommended testing protocol eliminates the false-positive results for those carrying the *RHD-CE-D* gene as well as the RhD pseudogene.⁶

The current body of literature supports the use of NIPD for the diagnosis of fetal RhD status. Four European clinical centers now offer this protocol as a standard of care. In the largest published trial from the Netherlands, 2359 samples from RhD-negative patients were obtained. The diagnostic accuracy of NIPD versus cord blood serology was 99.4%.¹⁵ Consequently, the use of NIPD for the analysis of the RhD genotype has virtually replaced invasive testing as the first-line method in these countries. In the

Consequently, the use of NIPD for the analysis of the RbD genotype has virtually replaced invasive testing as the first-line method in these countries.

United States, this testing remains in an experimental protocol. Prior to its widespread use, the accuracy of the various methods available must be evaluated critically. The use of automated methods that allow for reliable and inexpensive results needs to be further assessed. In the future, this system will most likely be applied to the multitude of antigen systems, including the Kell and Duffy systems, that are of clinical significance.⁹

Antenatal Diagnosis and Management of High-Risk Pregnancies

The feasibility of NIPD in the examination and management of pregnancy complications is currently under investigation. Researchers have begun assessing the use of cffDNA and fmRNA as surrogate biomarkers in a variety of pregnancy-related diseases. The source of these markers is known to be the placenta. As such, the possibility that these biomarkers may be used as indicators of placenta-mediated pregnancy complications has been raised. Placental markers isolated from the maternal circulation could possibly lead to the earlier diagnosis of a multitude of diseases. Included in this group of pregnancy complications would be preeclampsia (PEC), fetal growth restriction (FGR), and preterm labor (PTL).

Even though several sources of cffDNA in the maternal circulation have been proposed, the most likely source is the placental trophoblasts.¹⁶ It has been hypothesized that the high rate of trophoblastic apoptosis leads to the majority of the circulating cffDNA. Evidence to support this hypothesis includes the following¹⁷:

- The presence of circulating cffDNA prior to the development of fetoplacental circulation.¹⁸
- The rapid clearing of cffDNA with the removal of the placenta.¹⁹
- The presence of cffDNA in the maternal serum of anembryonic gestation.²⁰
- The absence of Y chromosome DNA in a mosaic Y-positive fetus with a concurrent placental 45X genotype.²¹

Since the discovery of circulating cffDNA, the idea that its presence and magnitude could be used in the prediction and management of pregnancy complications has been considered. Recently, it has been studied in placenta-mediated pregnancy complications such as PEC, FGR, and PTL. The hypothesis behind this finding is that the progressive placental dysfunction is manifested by increased rates of trophoblastic apoptosis. Therefore, any pregnancy complication associated with placental dysfunction may demonstrate increased circulating cffDNA.¹²

Preeclampsia

PEC is a disease of pregnancy that is thought to be mediated by a multitude of factors. Placental dysfunction has been implicated in the origin and progression of the disease.

Levels of cffDNA have definitively been shown to be increased in patients with PEC. Lo *et al.*²²

Levels of cffDNA have definitively been shown to be increased in patients with PEC.

demonstrated in their work a 5-fold increase in cffDNA in patients with PEC.

Levels of cffDNA have also been shown in small studies to increase prior to the onset of the disease itself.^{23,24} Similarly, the level of circulating cffDNA has been shown to be increased in the first trimester in patients who subsequently develop PEC.²⁵ This suggests the possibility of screening for PEC prior to the manifestation of maternal symptoms.

A direct proportional relationship between the severity of PEC and the level of cffDNA has also been shown in the recent literature.²⁶ All these findings must be validated on a large scale before the clinical implications can be fully appreciated.

Fetal Growth Restriction

The most common cause of FGR is thought to be placental dysfunction. The relationship between FGR and levels of cffDNA and fmRNA has begun to be examined. The fact that FGR may have multiple unrelated etiologies, however, has led to conflicting results in the literature. This controversy is compounded by the fact that most of the published studies include only very small numbers.

In an attempt to control for this problem and include only patients whose fetal FGR is related to placental dysfunction, Smid *et al.*²⁷ studied pregnancies with FGR and abnormal umbilical

Doppler waveforms. Their goal was to isolate those cases of FGR due to abnormal placental function. This study did in fact demonstrate increased levels of cffDNA in FGR fetuses in this subgroup.

Although this is promising, further research is necessary before the implementation of this strategy in clinical practice. Research now suggests a combination of NIPD and ultrasound evaluation as the best approach for pregnancies complicated by FGR.

Research now suggests a combination of NIPD and ultrasound evaluation as the best approach for pregnancies complicated by FGR.

Preterm Labor

The relationship between cffDNA and PTL remains poorly understood at this time. Farina *et al.*²⁸ recently demonstrated that a high concentration of cffDNA was associated with an increased risk of preterm delivery. The authors suggested the development of a cutoff to designate patients at high risk for preterm delivery. The authors conceded that one of the major limitations of this study was the inclusion of only 71 patients.

The major barrier in the development of NIPD testing is the multifactorial nature of PTL. The utility of cffDNA in the prediction and management of PTL is yet to be elucidated, and much work is still required.

The utility of cffDNA in the prediction and management of PTL is yet to be elucidated, and much work is still required.

CONCLUSION

The single belief that is common to the body of literature on NIPD is the need for further research. Our current understanding of NIPD remains in its infancy. Certain aspects of NIPD, such as fetal Rh and sex determination, are already in clinical use in certain countries. However, tests for other genetic abnormalities require further large-scale validation prior to their introduction into clinical practice. With the current pace of and interest in NIPD, significant advances in this field can be expected in the near future.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Schmorl G. *Pathologisch-Antomische Untersuchungen über Puerperal Eklampsie*. Leipzig, Germany: Verlag FCW Vogel; 1893.
- Lo YM, Corbetta N, Chamberlain PF, et al. Presence of fetal DNA in maternal plasma and serum. *Lancet* 1997; 350: 485–487.
- Poon LL, Leung TN, Lau TK, Lo YM. Presence of fetal RNA in maternal plasma. *Clin Chem* 2000; 46: 1832–1834.
- Hahn S, Zhong XY, Holzgreve W. Recent progress in non-invasive prenatal diagnosis. *Semin Fetal Neonatal Med* 2008; 13: 57–62.
- Bianchi DW, Simpson JL, Jackson LG, et al. Fetal gender and aneuploidy detection using fetal cells in maternal blood: analysis of NIFTY I data. National Institute of Child Health and Development Fetal Cell Isolation Study. *Prenat Diagn* 2002; 22: 609–615.
- Lo YM, Tsui NB, Chiu RW, et al. Plasma placental RNA allelic ratio permits non-invasive prenatal chromosomal aneuploidy detection. *Nat Med* 2007; 13: 218–223.
- Hahn S, Chitty LS. Non-invasive prenatal diagnosis: implications for antenatal diagnosis and the management of high-risk pregnancies. *Semin Fetal Neonatal Med* 2008; 13: 55–56.
- Chitty LS, van der Schoot CE, Hahn S, Avent ND. SAFE—the Special Non-Invasive Advances in Fetal and Neonatal Evaluation Network: aims and achievements. *Prenat Diagn* 2008; 28: 83–88.
- Fan HC, Quake SR. Detection of aneuploidy with digital polymerase chain reaction. *Anal Chem* 2007; 79: 7576–7579.
- Lo YM, Lun FM, Chan KC, et al. Digital PCR for the detection of fetal chromosomal aneuploidy. *Proc Natl Acad Sci U S A* 2007; 104: 13116–13121.
- Fan HC, Blumenfeld YJ, Chitkara U, et al. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci U S A* 2008; 105: 16266–16271.
- Lun FM, Tsui NB, Chan KC, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2008; 105: 19920–19925.
- Finning KM, Chitty LS. Non-invasive fetal sex determination: impact on clinical practice. *Semin Fetal Neonatal Med* 2008; 13: 69–75.
- van der Schoot CE, Hahn S, Chitty LS. Non-invasive prenatal diagnosis and determination of fetal Rh status. *Semin Fetal Neonatal Med* 2008; 13: 63–68.
- Van der Schoot CE, Soussan AA, Dee R, et al. Screening for foetal RhD-genotype by plasma PCR in all D-negative pregnant women is possible. *Vox Sang* 2004; 87(suppl): 9.
- Ishihara N, Matsuo H, Murakoshi H, et al. Increased apoptosis in syncytiotrophoblasts in human term

- placentas complicated by either PET or IUGR. *Am J Obstet Gynecol* 2002; 186: 158–166.
17. Alberry MS, Soothill PW. Non-invasive prenatal diagnosis: implications for antenatal diagnosis and management of high-risk pregnancies. *Semin Fetal Neonatal Med* 2008; 13: 84–90.
 18. Rijnders RJ, Van Der Luijt RB, Peters ED, et al. Earliest gestational age for fetal sexing in cell-free maternal serum. *Prenat Diagn* 2003; 23: 1042–1044.
 19. Lo YM, Zhang J, Leung TN, et al. Rapid clearance of fetal DNA from maternal plasma. *Am J Hum Genet* 1999; 64: 218–224.
 20. Alberry M, Maddocks D, Jones M, et al. Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblasts. *Prenat Diagn* 2007; 27: 415–418.
 21. Flori E, Doray B, Gautier E, et al. Circulating cell-free fetal DNA in maternal serum appears to originate from cyto and syncytio-trophoblastic cells: a case report. *Hum Reprod* 2004; 19: 723–724.
 22. Lo YM, Leung TN, Tein MS, et al. Quantitative abnormalities of fetal DNA in maternal serum in PET. *Clin Chem* 1999; 45: 184–188.
 23. Leung TN, Zhang J, Lau TK, et al. Increased maternal plasma fetal DNA concentrations in women who eventually developed PET. *Clin Chem* 2001; 47: 137–139.
 24. Zhong XY, Holzgreve W, Hahn S. The levels of circulating cell-free fetal DNA in maternal plasma are elevated prior to onset of PET. *Hypertens Pregnancy* 2002; 21: 77–83.
 25. Illanes S, Parra M, Serra R, et al. Increased free fetal DNA levels in the plasma of pregnant women in early pregnancy are observed in patients who subsequently develop preeclampsia and intrauterine growth restriction. *Prenat Diagn*. 2009; Oct 8. [Epub ahead of print]
 26. Swinkels DW, de Kok JB, Hendriks JC, et al. HELLP syndrome as a complication of PET in pregnant women increases the amount of cell-free fetal and maternal DNA in maternal plasma and serum. *Clin Chem* 2002; 48: 650–653.
 27. Smid M, Galbiati S, Lojcono A, et al. Correlation of fetal DNA levels in maternal plasma with Doppler status in pathological pregnancies. *Prenat Diagn* 2006; 26: 785–790.
 28. Farina A, LeShane ES, Romero R, et al. High levels of fetal cell-free DNA in maternal serum: a risk factor for spontaneous preterm labor. *Am J Obstet Gynecol* 2005; 193: 421–425.

Achieving a Safety Culture in Obstetrics

Erin DuPree, MD,¹ Loraine O'Neill, MPH, RN,² and Rebecca M. Anderson, MPH²

¹Department of Obstetrics, Gynecology, and Reproductive Science, Mount Sinai Medical Center, New York, NY

²Office for Excellence in Patient Care, Mount Sinai Medical Center, New York, NY

OUTLINE

BACKGROUND

SAFETY CULTURE

Just Culture
Reporting Culture
Learning Culture

PERINATAL SAFETY

Team Function and Communication
Evidence-Based Guidelines
Simulation

CONCLUSION

ABSTRACT

Preventable maternal and neonatal mortalities still occur, despite the wonders of today's technologically advanced healthcare system. Delivering high-quality, consistent care is the goal of every provider. Yet, obstetrical practice, in which unpredictable events and high-risk situations are the norm, is particularly vulnerable to medical errors. Obstetrics departments should be striving for a climate of patient safety, one that includes a just, reporting, and learning culture. This article discusses the various components of a safety culture as well as some of the advances that are being made in the field to improve the quality of care in obstetrics. *Mt Sinai J Med* 76:529–538, 2009. © 2009 Mount Sinai School of Medicine

Key Words: just culture, learning culture, obstetric patient safety, obstetric quality indicators, patient safety, reporting culture, safety culture.

"Healthy moms and healthy babies" is the unspoken motto of countless obstetrics departments nationwide, and yet despite the wonders of today's technologically advanced healthcare system, preventable perinatal deaths continue to occur. Many providers in obstetrics can enumerate several barriers that they encounter when delivering high-quality care and creating a trusting relationship with a patient, including liability issues, decreasing reimbursement, and the demands of highly sophisticated patients. The average obstetrician will face malpractice litigation 2.6 times in his or her career. From 1997 to 2003, the median payout in obstetrics for a brain-damaged infant was 2.6 million dollars. As a result, frustrated providers have limited their practices, avoided high-risk patients, "gone bare," or retired from obstetrics.^{1–3} This environment has created the perfect storm or rather the perfect opportunity by which to improve the overall quality of care by improvements in perinatal patient safety.

BACKGROUND

We have known since the 1980s that preventable injuries to patients are not rare. The Harvard Medical Practice Study found that 4% of patients in New York City had unintended injuries caused by their treatment.⁴ Only when the Institute of Medicine published *To Err is Human: Building a Safer Health System* did errors in medicine receive widespread publicity, alerting the public and government agencies to the issues of quality in medicine.⁵ A follow-up Institute of Medicine report, *Crossing the Quality Chasm: A New Health System for the 21st Century*, suggested that healthcare should be safe, patient-centered, timely, efficient, and equitable.⁶ In this construct, patient safety is a part of the larger quality concept. The Agency for Healthcare Research and Quality defines *patient safety* as "the absence of the potential for, or the occurrence of, health care associated injury to patients created by

Address Correspondence to:

Erin DuPree

Mount Sinai Medical Center
New York, NY

Email: erin.dupree@mssm.org

avoiding medical errors as well as taking action to prevent errors from causing injury.”⁷

Anesthesiologists became the pioneers of the patient safety movement in healthcare when a malpractice crisis in the 1970s led to the Anesthesia Patient Safety Foundation which reviewed closed claims and identified system improvements.^{8,9} The understanding and identification of preventable injuries led to a proactive approach by which solutions ultimately resulted in lower mortality rates and a reduction in malpractice premiums and lawsuits. The key lesson for obstetrics from the anesthesia experience is that safety and quality improved once the specialty internally addressed the issue of delivery of inconsistent, poor-quality care.

Medical errors are defined as mistakes made in the process of care that result in or have the potential to result in harm to patients.^{10,11} Mistakes include the failure of a planned action to be completed or the use of a wrong plan to achieve an aim: the result of an action that is taken (commission) or an action that should be taken but is not (omission).⁷ Errors may be made by any member of the healthcare team in any healthcare setting. A single error that passes through one of the organization's defenses may or may not lead to harm. A defense can include such things as properly functioning equipment, training of staff, or unit policies or procedures. Most often, harm happens to patients when errors line up and pass through weaknesses or “holes” in the defenses. This model is often referred to as the Swiss cheese model of organizational accidents (see Figure 1) because of the way in which errors need to breach multiple defenses in order to reach a patient. An underlying tenet of this model is that an organizational accident, such as a preventable injury in medicine, is most often the result of many contributing factors and local triggers, not a single action or event.

SAFETY CULTURE

A safety culture is the product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's health and safety management.¹² James Reason,¹³ an expert on human factors and systems analysis, suggested that to transform an average safety culture into an excellent one, an organization must recognize that a safe culture is dependent on several interlocking parts that can be engineered to some extent. A culture that continues to learn and improve (a learning culture) is dependent on a reporting culture, which in turn depends on establishing a just culture; see Figure 2 for an illustration of the interrelationships. A just culture is an atmosphere in which workers trust the leadership to treat them equitably. A safety culture built on trust is dependent on the sharing of information between workers and managers, including situations in which harm may have occurred but did not (ie, a near miss). Without a just culture, it is nearly impossible to establish an effective reporting culture or learning culture. It is the ability of an organization to engineer these aspects of its culture that leads to improved outcomes and a decrease in risk and preventable adverse events.

Certain industries, such as aviation and nuclear power, have achieved tremendous improvements in safety by designing systems that have effective barriers and defenses against human error. Organizations that have taken the lead and implemented such systems are known as high-reliability organizations.^{14,15} The aviation industry has acknowledged that error is an inevitable part of the human condition, especially in complex, stressful situations, and increasingly incorporates error management strategies to improve safety. Identified causes of error include fatigue, workload, and fear as well as cognitive

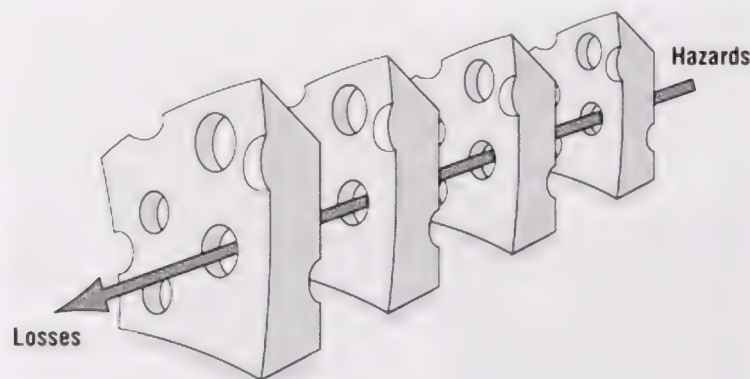


Fig 1. Swiss cheese model of organizational accidents.



Fig 2. Main subcomponents of a safe culture.

overload, poor interpersonal communications, imperfect information processing, and flawed decision making. Error management is based on understanding the nature and extent of error, changing the conditions that induce error, determining behaviors that prevent or mitigate error, and training personnel in their use.^{14–17} Fundamental to this paradigm is a shift from a punitive, naming-and-blaming approach to error, which still persists in much of healthcare. The creation of a safety culture is what will ultimately transform healthcare into a high-reliability industry in which errors and preventable adverse events are the exception and not the rule.

The creation of a safety culture is what will ultimately transform healthcare into a high-reliability industry in which errors and preventable adverse events are the exception and not the rule.

For obstetrics departments to nurture the development of a safety culture, it has to be acknowledged that errors occur. It is an organization's response to errors that determines whether a safety culture is supported. Therefore, it is imperative to create an environment in which staff feel secure reporting errors because they know that the information will be used in a nonpunitive manner for learning and improvement.

Just Culture

To ensure quality and to promote a culture of safety, a just culture must be in place, with trust between management and frontline staff.

To ensure quality and to promote a culture of safety, a just culture must be in place, with trust between management and frontline staff.

Once a trusting environment is created, staff will feel safe expressing concerns about safety and quality issues. It is within this environment that staff will perform at peak levels and also be able to admit when they, or others, are not delivering high-quality care consistently. Achieving the levels of trust needed for a just culture takes a tremendous amount of effort from all employees, regardless of title, within the organization.¹⁸

Healthcare leaders must address the problem of behaviors that threaten the performance of the healthcare team with the implementation of clearly defined professional behaviors. Any lack of clarity or accountability can result in diminished trust in leadership, which in turn results in increased suspicion, rumors, and defensiveness. This ultimately has a negative impact on the culture.¹⁹

The vast majority of the members of the healthcare team work under a code of professional conduct. However, intimidating and disruptive behaviors are not rare in healthcare, and they can erode other team members' high level of professionalism. These behaviors include overt actions such as verbal outbursts and physical threats as well as passive activities such as uncooperative attitudes during routine activities, a reluctance or refusal to answer questions or return phone calls or pages, condescending language or voice intonation, and impatience with questions.^{20–24} Intimidating and disruptive behaviors can create a culture that is hostile and leads to medical errors, poor patient satisfaction, increased litigation, and preventable adverse outcomes, all of which compromise patient safety.^{20–23,25–29}

Intimidating and disruptive behaviors can create a culture that is hostile and leads to medical errors, poor patient satisfaction, increased litigation, and preventable adverse outcomes, all of which compromise patient safety.

The creation of a just culture takes a commitment from leadership to change the way in which the

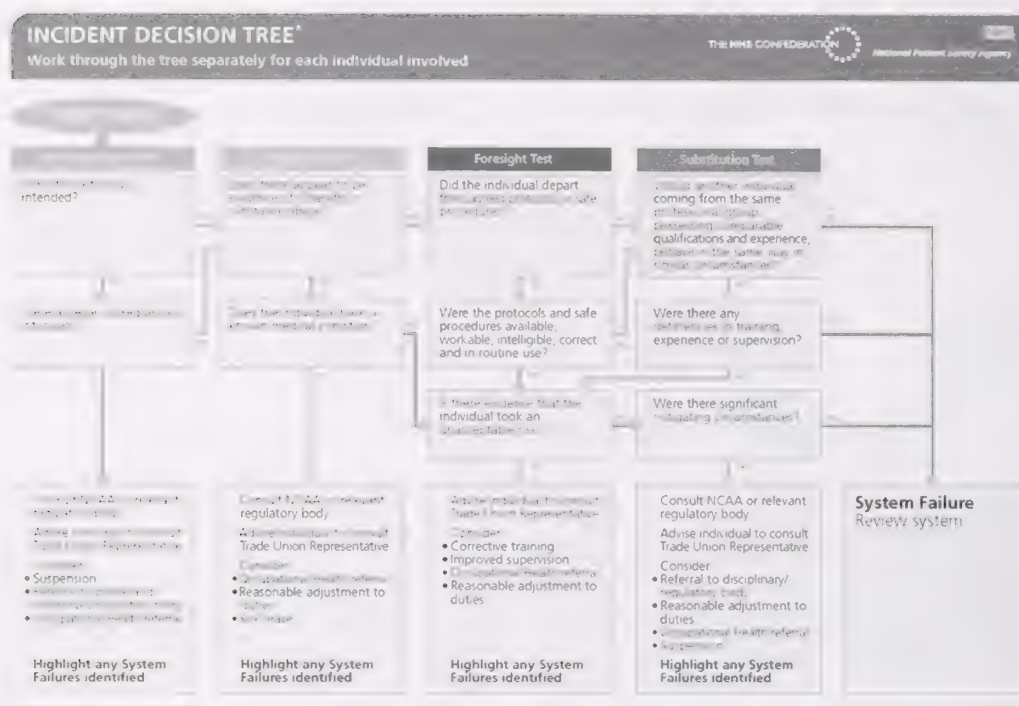


Fig 3. NHS Incident Decision Tree for responding to patient safety events. **Abbreviations:** NCAA, National Clinical Assessment Authority; NHS, National Health Service.

organization treats its staff. For many organizations, this may require a substantial transformation in the organizational culture. There are specific tools, such as codes of conduct and decision trees, that can be useful in fostering a change. The Joint Commission (TJC) has mandated that institutions have effective codes of conduct in place.³⁰ These codes are often a first step in the development of a clearly articulated vision of expected professionalism and behavior. Although various professional bodies have such codes, they are often limited in scope, addressing only the targeted profession, such as surgeons.^{31–34} Organizations should create codes of conduct that include all categories of staff and contain mechanisms for accountability. Consistent and sustained accountability leads to trust and patient safety and decreases overall liability.

Even though the majority of incident investigations will reveal many layers of system issues, leadership will confront situations in which there is a question of egregious individual error. The goal therefore is not to create an absolute no-blame culture because there are unsafe, egregious acts that cannot be tolerated and ultimately require disciplinary action. The United Kingdom Incident Decision Tree, as depicted in Figure 3, is a helpful algorithm used by the National Health Service to guide the investigation of incidents and make such determinations.^{35–37}

Reporting Culture

With the trust and mutual respect of a just culture in place, the foundation is laid for learning about the local system.

With the trust and mutual respect of a just culture in place, the foundation is laid for learning about the local system.

To support this process, members of the healthcare team must feel comfortable relating or reporting their errors and near misses. They must be assured that their reporting of events is tied to a responsive system that will take action. Staff will not become actively engaged in reporting if they do not see the effects of the reports incorporated into the daily work of the organization. Requiring team members to file incident reports is not always an easy task and can fail when the task is time-consuming. A fear of retribution and the trauma of reliving the experience can also impede reporting. Consistent and reliable reporting will happen only when frontline staff believe the reports will be used to drive change.³⁵

Mandatory systems of reporting can operate at an institutional and/or government level, and these typically focus on the identification of the most serious

adverse events and/or issues related to criminal activity, gross negligence, or professional misconduct. In addition to mandatory reporting, some institutions have implemented voluntary systems that are integral to learning and focus on reducing the occurrence of errors and improving patient safety by identifying system vulnerabilities. Reportable occurrences in a voluntary system should include any adverse event, near miss, or hazardous condition that could lead to an undesirable patient outcome. One of the best examples of a voluntary near-miss reporting system is the Patient Safety Reporting System. The Patient Safety Reporting System was jointly developed by the Department of Veterans Affairs and the National Aeronautics and Space Administration and invites everyone who works in a Department of Veterans Affairs facility to voluntarily report any events or concerns that involve patient safety.³⁸ It is designed to be complementary and external to the current internal adverse event reporting system at the Department of Veterans Affairs. It incorporates the following factors for a successful reporting program: indemnity against discipline (as far as it is reasonable), confidentiality, separation of the analysis of reports from bodies that have the authority to discipline, rapid feedback, and ease of making a report.³⁹

Another voluntary reporting system of reporting incidents is the New York State Safe Motherhood Initiative, by which the New York State Department of Health and the American College of Obstetricians and Gynecologists join together in efforts to understand maternal mortality. In this initiative, hospitals report maternal deaths, and an external group investigates the deaths. The information is collected, analyzed, and disseminated on a state level in an attempt to decrease preventable maternal deaths.⁴⁰ The pooling of data on a state level allows for the analysis of rare events.

Once reporting is established as an organizational priority, an approach to assessing what aspects of quality to measure can be selected. Quality indicators can be measured as structure, process, or outcome measures.^{41,42} Structure measures include the organizational context of care, such as the existence of an interdisciplinary practice committee and key clinical protocols for areas known to be associated with patient risk. Examples of structure measures in obstetrics include mandatory certification in electronic fetal monitoring and standard unit protocols for oxytocin, misoprostol, and magnesium sulfate administration.¹¹

Process measures evaluate how care is provided, such as adherence to evidence-based guidelines. Providers often feel more accountable for processes than outcomes.^{43,44} To be valid quality measures,

however, process measures must be causally linked to important outcomes (eg, the administration of prophylactic antibiotics to prevent a neonatal group B streptococcus infection).

Outcome measures are the most well-known and accessible type of quality indicator. These data are often available through administrative data and allow for large-scale reporting and trending. Outcome data, however, have some limitations with respect to validity.⁴⁴⁻⁴⁷ In order for certain outcomes to be used as quality indicators, risk adjustment must be performed to correct for differences among patients and hospitals.

At this time, adverse obstetrical events are not uniformly captured or reported, and this is reflected in the absence of validated obstetric-specific indicators of quality or patient safety. TJC and the Maryland Hospital Quality Indicator Project have used cesarean and vaginal birth after cesarean rates as measures of quality,^{48,49} but changes in practice (acceptance of elective cesarean delivery and safety of vaginal breech delivery) call the validity of these measures into question. The Agency for Healthcare Research and Quality Patient Safety Indicators have been put forth as measures to facilitate the identification of adverse events. The obstetrical indicators include such occurrences as third-degree and fourth-degree perineal tears and neonatal birth trauma.⁵⁰ Grobman *et al.*⁵¹ reviewed these and concluded that they are significantly influenced by both patient and hospital characteristics and therefore are not good obstetrical indicators of patient safety. Mann *et al.*⁵² developed an adverse outcome index, a weighted adverse outcome index, and a severity index from a set of outcome measures that were determined by consensus. They suggested that the measures can be used to benchmark ongoing care within and among organizations. In order for these to become useful on a national level, the measures will need to be risk-adjusted, as demonstrated by multiple nonobstetric conditions. However, before risk adjustment can be used, the factors that influence the adverse outcome index will need to be better understood.^{40,47}

To create a valid set of performance measures, obstetrics must examine its own unique care setting and conditions. It is imperative that professional societies, such as the American College of Obstetricians and Gynecologists and the Association of Women's Health, Obstetric, and Neonatal Nurses, agree on standard definitions and provide tools to caregivers to monitor performance. The importance of the measures must resonate with the frontline nurses and providers; otherwise, there will be no motivation to measure, report, and improve.

The importance of the measures must resonate with the frontline nurses and providers; otherwise, there will be no motivation to measure, report, and improve.

The reporting of adverse events and near misses is critical to improvement efforts, as the investigation, analysis, and prevalence of these incidents will continue to determine which areas should be targeted as quality indicators.

Learning Culture

Through heightened staff awareness, reporting, and data collection, adverse events can be turned into lessons learned. Human and systems failures can be reviewed through the use of root cause analysis or failure mode effective analysis. These are tools that can be employed to facilitate education and drill down on the details of an event to determine which defenses succeeded or failed. With a broad awareness of the systems' strengths and weaknesses, corrective action plans can be developed and implemented. The effectiveness of these plans in preventing repetitions of the same adverse event relies on a comprehensive interdisciplinary approach.

As a field, obstetrics is moving forward with many improvements in areas selected to improve patient safety. To date, these improvements have been focused on communication, teamwork, simulation of emergencies, and implementation of evidence-based guidelines and bundles of care. Having created a just environment and armed with knowledge from reporting, obstetrics departments are on their way to creating environments in which consistent and quality care is the norm.

PERINATAL SAFETY

Across the country, many obstetrics departments have adopted initiatives to improve patient safety. Several of these changes have been sponsored by malpractice carriers or healthcare companies, such as Ascension and Kaiser.^{53–58} Although these programs may be too new to readily identify specific changes in outcomes, some have reported reductions in severity of outcome.⁵⁹ The major focus in perinatal safety has been the integration of the science of high reliability with human factors research. These integrated approaches center on team function and communication, simulation, and evidence-based guidelines.

The major focus in perinatal safety has been the integration of the science of high reliability with human factors research. These integrated approaches center on team function and communication, simulation, and evidence-based guidelines.

Team Function and Communication

Through the work of Brian Sexton and others, culture and climate have been shown to affect outcome. The results of a survey measuring staff perception of a safety culture indicate that teamwork performance is directly correlated with the organizational culture.^{60,61} These surveys have been administered to over 600 hospital units: mostly intensive care units, emergency departments, and labor and delivery suites. Flattening of hierarchy, empowerment of junior staff, escalation processes, and effective, focused communication are techniques that can be implemented to improve an organizational culture. These changes, however, are often slow to implement and difficult to sustain. They are adaptations that demand a changing philosophy, the involvement of top-tier leadership, and transparency: in other words, the creation of a just culture.

Many labor and delivery units have begun to train their interdisciplinary teams in a formalized manner. There are several training format choices, but all revolve around the ability of effective resource management, the cohesion of a shared mental model (all on the same page; watching the same movie), and the delivery of timely, effective, and coordinated care. These tenets have been adopted from crew resource management and lessons learned in the cockpit.^{62–65} Thomas *et al.*⁶⁶ reported that team behaviors correlated with the overall rating of the quality of care, and this supports the rationale of studying team training in the delivery room further. These different training methods are best implemented in a culture in which trust and teamwork are already established.

Failures of communication have been shown to contribute to over 70% of adverse perinatal outcomes.⁶⁷ An example of a successful intervention to address communication is the adoption by both the American College of Obstetricians and Gynecologists and the Association of Women's Health, Obstetric, and Neonatal Nurses of the 1997 National Institute of Child Health and Human

Development nomenclature for the interpretation of electronic fetal monitoring.^{68,69} This implementation standardizes the tracing descriptions, and their use in daily practice has been supported by regulatory agencies, including TJC.⁶⁷ Many departments have instituted a common certification for all staff who interpret fetal monitor tracings.⁵³ In September 2008, the National Institute of Child Health and Human Development published an update, highlighting a 3-tiered system for the categorization of fetal heart tracings.⁷⁰

Strong efforts have been made to adopt a standardized handoff; for this, healthcare has again turned to the National Aeronautics and Space Administration and aviation. SBAR (situation, background, assessment, recommendation) has become widely adopted nationwide.⁷¹ Nursing communication tends to be more descriptive, and physician communication tends to be more action-oriented. SBAR can help to bridge the interdisciplinary gap and ensure that critical information is properly relayed and understood. In addition to SBAR, closed-loop communication can help to eliminate ambiguities in conversation. At the end of the communication, both parties need to be aware of the next steps and have a time frame: "I will be up to see the patient in 10 minutes." As mentioned earlier, accountability is a key component of a just culture.

Evidence-Based Guidelines

A care bundle is a group of evidence-based interventions related to a disease or care process that, when executed together, result in better outcomes than when they are implemented individually. The foremost example of evidence-based practices is the 100,000 Lives Saved program of the Institute for Healthcare Improvement, which demonstrated improvements in several clinical outcomes, including a significant reduction in surgical site infections.⁷²⁻⁷³ In relation to obstetrics, 2 perinatal care bundles are proposed by the Institute for Healthcare Improvement. These are the induction bundle, which requires (1) assessment of gestational age (ensuring that is greater than 39 weeks), (2) monitoring of the fetal heart rate for reassurance, (3) pelvic assessment and monitoring, and (4) management of hyperstimulation, and the augmentation bundle, which requires (1) an estimation of fetal weight, (2) monitoring of the fetal heart for reassurance, (3) pelvic assessment, and (4) monitoring and management of hyperstimulation.⁷⁴

Simulation

The Agency for Healthcare Research and Quality has stated the following:⁷⁵

Drills that are carefully planned can decrease medical errors by addressing unintended events that may result in injury to a patient arising from unintentional actions, mistakes in judgment, or inadequate plans of actions.

Drills are an opportunity for staff to learn to identify risk factors and prepare for obstetric emergencies. Simulation in perinatal settings has been recommended by TJC as a strategy to decrease preventable neonatal injury and death.⁶⁷ Optimal simulation exercises should involve all the key players, including obstetricians, anesthesiologists, pediatricians, labor nurses, midwives, and operating room staff. Simulation that is conducted in situ or that recreates the delivery suite is most productive in allowing all team members to practice their roles and to identify environmental barriers. The performance of shoulder dystocia drills facilitates the teamwork approach and the practice of associated maneuvers. Sites in both the United States and the United Kingdom have reported improvements in neonatal outcome and management of shoulder dystocia, including more detailed and accurate documentation.⁷⁶⁻⁸¹ Maternal hemorrhage, seizure, and emergent cesarean are other emergencies that can be recreated with simulation. A learning culture, in which team members can work together to identify strengths and weaknesses, positions an organization to be better prepared for adverse events and, most importantly, to best use the knowledge gained from the adverse event to prevent its occurrence in the future.

CONCLUSION

As the intimidation of liability continues to pressure the practice of obstetrics, the time has come for the perinatal safety movement. The goal in obstetrics should be the development of a patient safety culture: a just culture, a reporting culture, and a learning culture. Efforts should be directed at creating a culture of trust between those at the bedside and leadership. On a local level, institutions should initially ensure that their reporting systems for adverse events and near misses are robust. This local learning can then be used to engage in and inform regional, state, and national efforts for improvement. Armed with the important frontline knowledge of the delivery suite, obstetric healthcare

workers should then be empowered to take the lead in changing the environment of care; otherwise, regulators, lawyers, and administrators will dictate the care that is provided to patients.

As evidenced by the examples in this article, the perinatal safety movement is burgeoning. Obstetrics departments have begun to take a closer look within and to limit the ability of plaintiffs' attorneys to dictate practice standards. Researchers and regulatory agencies are actively searching for validated quality indicators. Institutions are introducing changes to the overall culture and targeting improvement efforts. Pearlman⁸² stated the following:

The effect of medical errors and unsafe systems of care has had a profound effect on the practice of obstetrics and gynecology ... We have the moral imperative as a specialty to fully engage in the identification of our own best practices, to advance safety research in obstetrics and gynecology, and to implement those practices which are best.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Wilson N, Strunk AL. Overview of the 2006 ACOG Survey on Professional Liability. Available at: <http://www.acog.org/departments/professionalliability/2006surveyNatI.pdf>. Accessed February 2009.
2. *Medical Malpractice: Verdicts, Settlements and Statistical Analysis, Updated Edition*. Horsham, PA: Jury Verdict Research; 2005; 12.
3. Studdert DM, Mello MM, Sage WM, et al. Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *JAMA* 2005; 293: 2609–2617.
4. Brennan T, Leape LL, Laire NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 1991; 324: 370–376.
5. Institute of Medicine. *To Err is Human: Building a Safer Health Care System*. Washington, DC: National Academy Press; 2000.
6. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
7. *AHRQ's Patient Safety Initiative: Building Foundations, Reducing Risk: Interim Report to the Senate Committee on Appropriations*. AHRQ publication 04-RG005, December 2003, AHRQ, Rockville, MD. Available at: <http://www.ahrq.gov/qual/pscongrpt>. Accessed March 2009.
8. Cheney FW. The American Society of Anesthesiologists Closed Claims Project: what have we learned, how has it affected practice, and how will it affect practice in the future? *Anesthesiology* 1999; 91: 552–556.
9. Lagasse RS. Anesthesia safety: model or myth? A review of the published literature and analysis of current original data. *Anesthesiology* 2002; 97: 1609–1617.
10. Reason JT. *Human Error*. Cambridge, England: Cambridge University Press; 1990.
11. Leape LL. Error in medicine. *JAMA* 1994; 272: 1851–1857.
12. ACSNI Study Group on Human Factors. *Third Report: Organizing for Safety*. London, United Kingdom: Health and Safety Commission; 1993.
13. Reason J. *Managing the Risks of Organizational Accidents*. Aldershot, United Kingdom: Ashgate Publishing; 1997.
14. McKeon LM, Oswaks JD, Cunningham PD. Safeguarding patients: complexity science, high reliability organizations, and implications for team training in healthcare. *Clin Nurse Spec* 2006; 20: 298–304; quiz 305–306.
15. Weick KE, Sutcliffe KM. *Managing the Unexpected: Assuring High Performance in an Age of Complexity*. San Francisco, CA: Jossey-Bass; 2001.
16. Baker DP, Day R, Salas E. Teamwork as an essential component of high-reliability organizations. *Health Serv Res* 2006; 41(pt 2): 1576–1598.
17. Helmreich RL. On error management: lessons from aviation. *BMJ* 2000; 320: 781–785.
18. Frankel AS, Leonard MW, Denham CR. Fair and just culture, team behavior, and leadership engagement: tools to achieve high reliability. *Health Serv Res* 2006; 41: 1690–1709.
19. MacDavitt K, Chou SS, Stone PW. Organizational climate and health care outcomes. *Jt Comm J Qual Patient Saf* 2007; 33(suppl): 45–56.
20. Rosenstein AH, O'Daniel M. Disruptive behavior and clinical outcomes: perceptions of nurses and physicians. *Am J Nurs* 2005; 105: 54–64.
21. Institute for Safe Medication Practices. Survey on workplace intimidation 2003. Available at: <https://ismep.org/Survey/surveyresults/Survey0311.asp>. Accessed April 2008.
22. Gerardi D. The emerging culture of health care: improving end-of-life care through collaboration and conflict engagement among health care professionals. *Ohio State J Dispute Resolution* 2007; 23: 105–142.
23. Weber DO. Poll results: doctors' disruptive behavior disturbs physician leaders. *Physician Exec* 2004; 30: 6–14.
24. Leape LL, Fromson JA. Problem doctors: is there a system-level solution? *Ann Intern Med* 2006; 144: 107–155.
25. Morrissey J. Encyclopedia of errors: growing database of medication errors allows hospitals to compare their track records with facilities nationwide in a nonpunitive setting. *Mod Healthc* 2003; 33: 40–42.
26. Gerardi D. Effective strategies for addressing "disruptive" behavior: moving from avoidance to engagement. Paper presented at: Medical Group Management Association Webcast; 2007.
27. Gerardi D. Creating cultures of engagement: effective strategies for addressing conflict and "disruptive" behavior. Available at: [http://law.creighton.edu/wernerInstitute/healthcareconference/speaker_materials/disruptive behavior presentation - Debra Gerardi.pdf](http://law.creighton.edu/wernerInstitute/healthcareconference/speaker_materials/disruptive%20behavior%20presentation%20-%20Debra%20Gerardi.pdf). Accessed March 2009.
28. Neff KE. *Understanding and managing physicians with disruptive behavior*. In: Ransom SB, Pinsky WW,

- Tropman JE (eds.) *Enhancing Physician Performance: Advanced Principles of Medical Management*. Tampa, FL: American College of Healthcare Executives; 2000; 45–72.
29. Hickson GB. A complementary approach to promoting professionalism: identifying, measuring, and addressing unprofessional behaviors. *Acad Med* 2007; 82: 1040–1048.
 30. Joint Commission on Accreditation of Healthcare Organizations. Sentinel Event Alert 40, 2008. Available at: http://www.jointcommission.org/SentinelEvents/SentinelEventAlert/sea_30.htm. Accessed March 2009.
 31. Committee on Patient Safety and Quality Improvement of the American College of Obstetricians and Gynecologists. ACOG Committee Opinion. Number 366 May 2007. Disruptive behavior. *Obstet Gynecol* 2007; 109: 1261–1262.
 32. American College of Obstetricians and Gynecologists. Code of professional ethics, January 2008. Available at: http://www.acog.org/from_home/acogcode.pdf. Accessed March 2009.
 33. American College of Surgeons. Code of professional conduct, November 2005. Available at: <http://www.facs.org/memberservices/codeofconduct.html>. Accessed March 2009.
 34. American Medical Association. Principles of medical ethics. Available at: <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/ama-code-medical-ethics/principles-medical-ethics.shtml>. Accessed March 2009.
 35. Reason JT. *Managing the Risks of Organizational Accidents*. Aldershot, United Kingdom: Ashgate Publishing; 2003; 191–222.
 36. Reason J, Hobbs A. *Managing Maintenance Error: A Practical Guide*. Aldershot, United Kingdom: Ashgate Publishing; 2003.
 37. National Patient Safety Agency. Incident Decision Tree. Available at: <http://www.npsa.nhs.uk/nrls/improvingpatientsafety/patient-safety-tools-and-guidance/incidentdecisiontree/>. Accessed March 2009.
 38. Patient Safety Reporting System. Available at: <http://www.psr.sarc.nasa.gov/flashsite/programoverview/index.html>. Accessed March 2009.
 39. Leape LL. Reporting of adverse events. *N Engl J Med* 2002; 347: 1633–1638.
 40. American College of Obstetrics and Gynecology. NYS Safe Motherhood Initiative. Available at: http://www.acog.org/acog.districts/dist_notice.cfm?recno=1&bulletin=1551. Accessed March 2009.
 41. Simpson KR. Measuring perinatal patient safety: review of current methods. *J Obstet Gynecol Neonatal Nurs* 2006; 35: 432–442.
 42. Donabedian A, Bashshur R. *An Introduction to Quality Assurance in Health Care*. New York, NY: Oxford University Press; 2003.
 43. Rubin H, Pronovost P, Diette G. From a process of care to a measure: the development and testing of a quality indicator. *Int J Qual Health Care* 2001; 13: 489–496.
 44. Rubin H, Pronovost P, Diette G. The advantages and disadvantages of process-based measures of health care quality. *Int J Qual Health Care* 2001; 13: 469–474.
 45. Chassin, MR, Hannan E, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. *N Engl J Med* 1996; 334: 394–398.
 46. Myers ER, Steege JF. Risk adjustment for complications of hysterectomy: limitations of routinely collected administrative data. *Am J Obstet Gynecol* 1999; 181: 567–575.
 47. Parker JP, Li Z, Damberg CL, et al. Administrative versus clinical data for coronary artery bypass graft surgery report cards. *Med Care* 2006; 44: 687–695.
 48. Joint Commission on Accreditation of Healthcare Organizations. A comprehensive review of development and testing for national implementation of hospital core measures. Available at: <http://www.jointcommission.org/NR/rdonlyres/48DFC95A-9C05-4A44-AB05-1769D5253014/0/AComprehensiveReviewofDevelopmentforCoreMeasures.pdf>. Accessed March 2009.
 49. Maryland Hospital Quality Indicator Project. Quality indicator project acute care measures. Available at: http://www.qiproject.org/pdf/Acute_Care_Indicators.pdf. Accessed March 2009.
 50. National Quality Measures Clearinghouse. Available at: <http://www.qualitymeasures.ahrq.gov/browse/brosenode.aspx?node=14786&type=1>. Accessed February 2009.
 51. Grobman WA, Feinglass J, Murthy S, et al. Are the agency for healthcare research and quality obstetric trauma indicators valid measures of hospital safety? *Am J Obstet Gynecol* 2006; 195: 868–874.
 52. Mann S, Pratt S, Gluck P, et al. Assessing quality in obstetrical care: development of standardized measures. *Jt Comm J Qual Patient Saf* 2006; 32: 497–505.
 53. Will SB, Hennessee KP, Jacobs LS, et al. The patient safety nurse: a new role to promote safe care for mothers and babies. *J Obstet Gynecol Neonatal Nurs* 2006; 35: 417–423.
 54. Controlled Risk Insurance Company/Risk Management Foundation. Available at: <http://www.rm.f.harvard.edu/high-risk-areas/obstetrics/ob-risk-reduction-program/general-information.aspx>. Accessed March 2009.
 55. McFerran S, Nunes J, Pucci D, Zuniga A. Perinatal patient safety project: a multicenter approach to improve performance reliability at Kaiser Permanente. *J Perinat Neonatal Nurs* 2005; 19: 37–45.
 56. Mazza F, Kitchens J, Kerr S, et al. Eliminating birth trauma at Ascencion Health. *Jt Comm J Qual Patient Saf* 2007; 33: 15–24.
 57. Mazza F, Kitchens J, Akin M, et al. The road to zero preventable birth injuries. *Jt Comm J Qual Patient Saf* 2008; 34: 201–205.
 58. Knox GE, Simpson KR, Townsend KE. High-reliability perinatal units: further observations and a suggested plan for action. *ASHRM J* 2003; 23: 17–21.
 59. Pettker CM, Thung SF, Norwitz ER, et al. Impact of a comprehensive strategy to reduce obstetric adverse events. *Am J Obstet Gynecol* 2009; 200(492): e1–e8.
 60. Sexton JB. Effecting change: from attitudes to behaviors to outcomes. Paper presented at: the 2004 Annual Meeting of the Society for Healthcare Epidemiology of America; April 2004; Philadelphia, PA.
 61. Mann S, Marcus R, Sachs B. Lessons from the cockpit: how team training can reduce errors on L&D. *Contemp OB/GYN* 2006; 51: 34–45.
 62. McConaughy E. Crew resource management in healthcare: the evolution of teamwork training and MedTeams. *J Perinat Neonatal Nurs* 2008; 22: 96–104.

63. Haller G, Garnerin P, Morales MA, et al. Effect of crew resource management training in a multidisciplinary obstetrical setting. *Qual Health Care* 2008; 20: 254–263.
64. Nielsen P, Mann S. Team function in obstetrics to reduce errors and improve outcomes. *Obstet Gynecol Clin North Am* 2008; 35: 81–95, ix.
65. Nielsen PE, Goldman MB, Mann S, et al. Effects of teamwork training on adverse outcomes and process of care in labor and delivery: a randomized controlled trial. *Obstet Gynecol* 2007; 109: 48–55.
66. Thomas EJ, Sexton JB, Lasky RE, et al. Teamwork and quality during neonatal care in the delivery room. *J Perinatol* 2006; 26: 163–169.
67. Joint Commission on Accreditation of Healthcare Organizations. Sentinel Event Alert 30, 2004. Available at: http://www.jointcommission.org/SentinelEvents/SentinelEventAlert/sea_30.htm. Accessed March 2009.
68. Electronic fetal heart rate monitoring: research guidelines for interpretation. National Institute of Child Health and Human Development Research Planning Workshop. *Am J Obstet Gynecol* 1997; 177: 1385–1390.
69. American College of Obstetrics and Gynecology. Practice bulletin no. 70. Intrapartum fetal heart rate monitoring. *Obstet Gynecol* 2005; 106: 1453–1460.
70. Macones GA, Hankins GD, Spong CY, et al. 2008 National Institute of Child Health and Human Development Workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines. *Obstet Gynecol* 2008; 112: 661–666.
71. Haig KM, Sutton S, Whittington J. SBAR: a shared mental model for improving communication between clinicians. *Jt Comm J Qual Patient Saf* 2006; 32: 167–175.
72. Odom-Forren J. Preventing surgical site infections. *Nursing* 2006; 36: 58–63.
73. Resar R, Pronovost P, Haraden C, et al. Using a bundle approach to improve ventilator care processes and reduce ventilator-associated pneumonia. *Jt Comm J Qual Patient Saf* 2005; 33: 243–248.
74. Cherouney PH, Federico FA, Haraden C, et al. Idealized design of perinatal care. Available at: <http://www.ihl.org/IHI/Results/WhitePapers/IdealizedDesignofPerinatalCareWhitePaper.htm>. Accessed February 2009.
75. Shojania KG, Duncan BW, McDonald KM. Making Healthcare Safer #58: A Critical Analysis of Patient Safety Practices. Rockville, Md: Agency for Healthcare Research and Quality, 2001.
76. Draycott T, Sibanda T, Owen L, et al. Does training in obstetric emergencies improve neonatal outcome? *BJOG* 2006; 113: 177–182.
77. Goffman D, Heo H, Chazotte C, et al. Using simulation training to improve shoulder dystocia documentation. *Obstet Gynecol* 2008; 112: 1284–1287.
78. Draycott TJ, Crofts JF, Ash JP, et al. Improving neonatal outcome through practical shoulder dystocia training. *Obstet Gynecol* 2008; 112: 14–20.
79. Crofts JF, Bartlett C, Ellis D, et al. Training for shoulder dystocia: a trial of simulation using low-fidelity and high-fidelity mannequins. *Obstet Gynecol* 2006; 108: 1477–1485.
80. Deering S, Poggi S, Macedonia C, et al. Improving resident competency in the management of shoulder dystocia with simulation training. *Obstet Gynecol* 2004; 103: 1224–1228.
81. Goffman D, Heo H, Pardani S, et al. Improving shoulder dystocia management among resident and attending physicians using simulations. *Am J Obstet Gynecol* 2008; 199(294): e1–e5.
82. Pearlman MD. Patient safety in obstetrics and gynecology: an agenda for the future. *Obstet Gynecol* 2006; 108: 1266–1271.

Obesity and Pregnancy: Implications and Management Strategies for Providers

Taraneh Shirazian, MD, and Sreekala Raghavan, BS

Mount Sinai School of Medicine, New York, NY

OUTLINE

METHODS

MATERNAL COMPLICATIONS

FETAL COMPLICATIONS

INTRAPARTUM AND

POSTPARTUM COMPLICATIONS

PRENATAL WEIGHT GAIN GUIDELINES

INTERVENTIONAL APPROACHES

FUTURE DIRECTIONS

ABSTRACT

Obesity in pregnancy (pregravid body mass index ≥ 30) has been linked to several adverse pregnancy outcomes, including spontaneous abortion, preeclampsia, gestational diabetes, fetal macrosomia, cesarean delivery, and wound complications post-cesarean section. Intrapartum and postpartum management of obese gravidas requires multidisciplinary consultations between obstetricians, anesthesiologists, nurses, and pediatricians in order to improve the pregnancy outcomes of the mother and neonate. The American College of Obstetricians and Gynecologists currently supports risk-reducing strategies for obese pregnant patients, including limiting weight gain to 15 lb (standardized by the Institute of Medicine). Interventions to reduce gestational weight gain may be important modifiable risk factors for maternal and fetal perinatal complications. Interventions have targeted modifications of diet and exercise with educational methods such as radio broadcasts, pamphlets, and counseling. Interventions

have also focused on motivational methods, such as individual and group classes, and have been implemented both before conception and immediately after birth. Effective interventions appear to be individualized in approach, but there is a lack of data to support any specific model. Prospective interventional studies are needed to demonstrate the benefits of weight limitation on pregnancy outcomes. *Mt Sinai J Med* 76:539–545, 2009. © 2009 Mount Sinai School of Medicine

Key Words: obesity, pregnancy, risk-reducing strategies, targeted interventions.

Obesity is a national epidemic that has significant obstetric implications. The World Health Organization and the Centers for Disease Control and Prevention define normal weight as a body mass index (BMI) of 18.5 to 24.9 kg/cm², overweight as a BMI of 25 to 29.9 kg/cm², and obesity as a BMI ≥ 30 kg/cm².¹ One-third of Americans are obese, with one-half of women between the ages of 20 to 39 being defined as overweight or obese.² This is a sharp rise in comparison with the US prevalence rate a little over a decade ago: 22.9% in 1988–1994.³ Rates of obesity also differ by ethnicity: 49% of non-Hispanic white women versus 70% of non-Hispanic black women of child-bearing years are overweight or obese.⁴ The National Health and Nutrition Examination Survey (1999–2002), a survey of self-reported weight, indicates that 26% of reproductive-age women (20–39 years old) identify themselves as overweight, and 29% identify themselves as obese.⁵

Both prepregnancy weight and gestational weight gain among all weight groups have been sources of clinical investigations. From 1999 to 2003, an overall increase in the prepregnancy BMI of the total delivery population in New York was noted with an 11% increase in the overweight group as well as an 8% increase in an obese subgroup in comparison with healthy weight controls.⁶ The effect of gestational

Address Correspondence to:

Taraneh Shirazian

Mount Sinai School of Medicine
New York, NY

Email: taraneh.shirazian@mssm.edu

weight gain on pregnancy outcome is affected by a woman's prepregnancy BMI. A pregnancy weight gain of 6.7 to 11.2 kg (15–25 lb) in overweight and obese women and a pregnancy weight gain less than 6.7 kg (15 lb) in morbidly obese women are associated with a reduction in the risk of adverse outcomes.⁷

Retrospective data from birth registries have shown better outcomes with less gestational weight gain in all weight categories. In patients with a prepregnancy BMI ≥ 30 , gestational weight gain may be most important. The optimal pregnancy outcome with the minimum risk of preeclampsia, cesarean delivery, and large-for-gestational-age infants was associated with limited gestational weight gain per BMI category: a weight gain of 10 to 25 lb for women with class I obesity (BMI = 30–34.9), a weight gain of 0 to 9 lb for women with class II obesity (BMI = 35–39.9), and a weight loss of 0 to 9 lb for women with class III or extreme obesity (BMI = 40+).⁸ Interventions to limit gestational weight gain in obese patients have demonstrated a reduction in maternal and fetal complications.

This review will serve as a broad overview of obesity and pregnancy for the obstetrician/gynecologist. It will attempt to prepare the provider for management of this at-risk population. Maternal and fetal complications of obesity in pregnancy, intra-partum and postpartum issues for obese pregnant patients, and interventional programs developed to minimize these complications will be discussed.

METHODS

A literature search was conducted with PubMed/Medline for articles related to obesity, as defined by World Health Organization criteria, and pregnancy. Thirty-nine articles were reviewed on the basis of their relevancy to topics of maternal, fetal, and delivery complications related to obesity as well as interventions designed for obese pregnancy patients.

MATERNAL COMPLICATIONS

An increase in prepregnancy weight has been associated with multiple medical and obstetric problems, including increased risks of hypertension, preeclampsia, pregestational diabetes mellitus (DM), and gestational DM. In comparison with matched controls,

An increase in prepregnancy weight has been associated with multiple medical and obstetric problems, including increased risks of hypertension, preeclampsia, pregestational diabetes mellitus (DM), and gestational DM.

patients who enter pregnancy obese have been shown to be 2.5 times more likely to develop gestational hypertension, and morbidly obese patients have been shown to be 3.2 times more likely to develop gestational hypertension. Similarly, preeclampsia has been shown to be 1.6 and 3.3 times more likely in obese and morbidly obese patients, respectively.⁹ Obese and overweight women had higher risks of recurrent preeclampsia (19.3% and 14.2%) in comparison with women with a normal BMI (11.2%).¹⁰

A recent meta-analysis of studies examining the risk of gestational DM in overweight and obese women found odds ratios of 3.56 and 8.56 for gestational DM in obese and morbidly obese women, respectively, in comparison with normal-weight women.¹¹ Although most women with gestational DM return to a state of normal insulin response after delivery, approximately 5% to 10% of these women remain diabetic after delivery, and 20% to 50% will develop type 2 diabetes in their lifetime.¹¹ In addition, women with poorly controlled gestational DM at the time of delivery are at higher risk for neonatal respiratory distress, premature rupture of membranes, preterm delivery, cesarean section, macrosomia, and intrauterine fetal demise in the last 4 to 6 weeks of pregnancy.^{12–14} The obstetrician/gynecologist should be aware of the increased need for surveillance of obese pregnant women for early detection of these complications.

FETAL COMPLICATIONS

Obesity in pregnancy not only affects the mother but also leads to multiple fetal complications. Obese

Obesity in pregnancy not only affects the mother but also leads to multiple fetal complications.

pregnancy BMI has been linked to a higher risk of fetal neural tube defects, aneuploidy, spontaneous

miscarriage, intrauterine fetal demise, macrosomia, shoulder dystocia, neonatal respiratory distress, and preterm delivery. Some studies have found a relative risk ratio of 1.70 in obese women ($\text{BMI} \geq 30 \text{ kg/cm}^2$) and a relative risk ratio of 3.11 in severely obese women ($\text{BMI} \geq 38 \text{ kg/cm}^2$) for aneuploidy in comparison with normal-weight controls.¹⁵ Serum screening for aneuploidy and neural tube defects is usually adjusted for maternal weight; however, the test parameters may be less accurate at extremes of weight. In a meta-analysis by Chu *et al.*,¹⁶ the odds ratio of obese women having a stillbirth was 2.07 in comparison with normal-weight pregnant women. Morbidly obese women ($\text{BMI} \geq 40$) may not perceive normal fetal movement as their slimmer counterparts; therefore, they may not be able to detect a decrease in fetal movement during pregnancy and may be at increased risk for undetected stillbirth. There is an increased false-negative rate in detecting fetal anomalies by ultrasound in obese women in comparison with nonobese women because of difficulties in visualizing defects.¹⁷ In fact, a mid-trimester scan is difficult because of the nonvisualization of specific anatomical structures such as the heart, cerebellum, lips, kidneys, and skeletal structures,¹⁸ and some have recommended the use of a transvaginal scan at 12 to 15 weeks in order to maximize visualization.

INTRAPARTUM AND POSTPARTUM COMPLICATIONS

Obesity significantly affects obstetric care. Physicians often have difficulty estimating fetal growth by physical examination, obtaining fetal heart tones, and performing contraction monitoring in labor. They are also significantly hindered in performing emergency cesarean section because of central adipose distribution, and these patients are more likely to undergo and fail induction of labor and to require a cesarean section. Obese women are more likely to have

Obesity significantly affects obstetric care. Physicians often have difficulty estimating fetal growth by physical examination, obtaining fetal heart tones, and performing contraction monitoring in labor. They are also significantly hindered in performing emergency cesarean section.

induction of labor in comparison with normal-weight women, although indications are less clear but may be related to cephalopelvic disproportion.^{19,20} Prepregnancy BMI has also been directly associated with failed trial of labor after prior cesarean delivery: the rate rises from 15.2% in normal-weight women to 39.3% in morbidly obese women, and the risk of rupture and scar dehiscence increases from 0.9% to 2.1%.²¹ Scheduled elective cesarean deliveries in morbidly obese women ($\text{BMI} \geq 40$) require advanced planning in a variety of disciplines. Special equipment must be secured for these patients, including wider operating tables, longer surgical instruments, wider surgical instruments such as retractors, and additional surgical assistants to provide visualization for the surgeons. This may pose a risk in the event of an emergent cesarean section if these tools are not readily available. There is also a greater risk of operative complications, including higher rates of excessive blood loss, longer operative times, cesarean wound infections, and endometritis in this subgroup of women.²² In women with excessive central adipose deposition, a periumbilical vertical incision may be needed in order to access the lower uterine segment for delivery of the fetus.

Anesthetic management of obese patients is also challenging; epidural and spinal placement is more difficult than in normal-weight women and results in repeated attempts and use of longer needles.²³ Further complications include respiratory failure from difficult intubations, with a greater rate of failure of intubation in these women.²⁴ Sleep apnea is a major risk factor in the anesthetic management of this group, and so is the need for increased doses of anesthetic agents.²⁶ Obese women may find it difficult to breathe in a supine position because of the weight of the breasts and abdomen.

Breastfeeding differences have also been found postpartum in comparisons of obese women and women of normal weight. A systematic review in 2007 examining breastfeeding in obese women showed that obese women have lower rates of breastfeeding intention, are less likely to initiate breastfeeding, and, if breastfeeding is initiated, are less likely to continue postpartum in comparison with normal-weight women.²⁵ Encouragement of breastfeeding may aid in promoting postpartum weight loss. A lactation consultant can be helpful in providing educational resources and support for women with large breasts who are motivated to nurse their infants.

In summary, intrapartum management of the fetus is challenging in these women, cesarean delivery carries pronounced risk to the mother and fetus, and postpartum issues related to obesity should be anticipated. Prevention of these complications

should begin during prenatal management through the provision of weight gain guidelines for these women.

PRENATAL WEIGHT GAIN GUIDELINES

The Institute of Medicine (IOM) created guidelines in 1990 for prenatal weight gain.²⁶ These guidelines were developed originally in an attempt to reduce the incidence of low-birth-weight neonates by encouraging appropriate maternal gestational weight gain. These guidelines were recently updated in 2009. New IOM guidelines recommend a total weight gain of 5 to 9 kg in obese women.²⁷ The current American College of Obstetricians and Gynecologists recommendations for prenatal weight gain are still based on the 1990 IOM guidelines: 11.5 to 16 kg for normal-weight patients (BMI = 18.5–24.9 kg/m²), 7 to 11.5 kg for overweight patients (BMI = 25–29.9 kg/m²), and at least 6 kg for obese patients (BMI ≥ 30 kg/m²).²⁸ Patients are categorized into weight classes on the basis of their prepregnancy weight. Weight gain within the ranges of these guidelines is associated with appropriate birth weight and fewer obstetric complications than maternal weight gain that exceeds or does not meet recommendations.²⁸ However, a 1999 study by Schieve *et al.*²⁹ showed that maternal gestational weight gain averages exceeded the guidelines in both the overweight and obese categories according to data recorded by the Centers for Disease Control and Prevention Pregnancy Nutrition Surveillance System. Knowledge and appropriate patient education regarding these guidelines are ways of preventing or reducing complications. In fact, although surveyed obstetricians/gynecologists are aware of the complications associated with obesity and pregnancy, many feel their training in weight management is inadequate.³⁰

The American College of Obstetricians and Gynecologists has provided other recommendations in order to guide obstetricians/gynecologists in taking care of this at-risk population. These recommendations include preconception counseling, educating women on the possible complications, early screening for gestational diabetes, and anesthesiology consultation.²⁸ Given the innumerable challenges and complications associated with obesity and pregnancy, it is imperative not only that providers counsel and educate their patients but that health care centers offer programs and preventive measures to curb the growing number of obese pregnant patients and to minimize gestational weight gain in obese patients.

Given the innumerable challenges and complications associated with obesity and pregnancy, it is imperative not only that providers counsel and educate their patients but that health care centers offer programs and preventive measures to curb the growing and to minimize gestational weight gain in obese patients.

INTERVENTIONAL APPROACHES

The US Preventive Services Task Force found that the most successful nonsurgical approaches for obesity management are highly intensive weight-focused counseling sessions about diet, exercise, and behavioral interventions. Pregnancy provides a captive

The US Preventive Services Task Force found that the most successful nonsurgical approaches for obesity management are highly intensive weight-focused counseling sessions about diet, exercise, and behavioral interventions.

audience for intervention, as many newly pregnant women are likely to modify their behavior in order to promote the health of the growing fetus. This has been studied primarily in smoking cessation interventions during pregnancy, when the highest proportion of women stop smoking.³¹ Consequently, pregnancy may be a unique time to intervene and to both minimize gestational weight gain on the basis of known guidelines and to educate women about lifelong weight management.

A recent comprehensive analysis of interventions during pregnancy targeting gestational weight gain reported a small number of prospective interventions for pregnancy and postpartum weight management.³² Six prospective interventions have been published to date. Two of those studies have published design protocols, but they are still in progress with no published results. These studies may help guide health care and community centers in developing programs to prevent gestational weight gain in obese patients.

The earliest of the interventional studies was conducted by Gray-Donald *et al.*³³ in the Canadian Cree population in 2000. The population has a high prevalence of gestational DM and type 2 DM. It was a prospective cohort study in which the intervention was offered to 112 pregnant women defined as obese (BMI ≥ 30). The intervention targeted diet modification, exercise with radio broadcasts, educational pamphlets, supermarket tours and cooking classes as nutritional activities, walking groups as a physical activity, and one-on-one counseling on individual diet choices. The intervention showed no significant difference in the primary endpoints of weight gain during the second half of pregnancy, plasma glucose levels between 24 and 30 weeks, birth weights, and maternal weight at 6 weeks postpartum. They concluded that interventions achieving dietary and physical activity modification remained inconclusive.

Polley *et al.*³⁴ in 2002 implemented a program in which pregnant women with a BMI >19.8 were assigned to either a behavior modification group or a control group. In the behavioral group, women were provided education on appropriate weight gain in pregnancy and instructions on healthy eating habits and exercise. A stepped care approach was used, in which each woman was given increasingly structured behavioral goals at each visit if her weight continued to exceed the recommended values. Patients receiving positive feedback as part of the intervention were found to adhere to the recommended IOM weight gain guidelines. However, this impact on excessive gestational weight gain was largely in normal-weight women and did not show a significant impact in overweight or obese women.

In another study by Olson *et al.*³⁵ in 2004, the intervention consisted of educational material mailed to patients with normal and overweight BMIs early in pregnancy, who were compared to historical controls that did not receive mailings. Gestational weight gain was monitored throughout pregnancy and for 1 year postpartum. The findings suggested that overweight patients in the low-income subgroup were less likely to retain weight at 1 year when educated on the effects of obesity.³⁵

Stotland *et al.*³⁶ published the most recent interventional study results in 2005. The study involved 4 telephone surveys during weeks 20, 24 to 28, and 32 to 36 of gestation and at 8 to 12 weeks postpartum. The survey questioned a patient's understanding of appropriate gestational weight gain and provider recommendations on suggested weight gain. The study demonstrated that women had widely differing beliefs about appropriate gestational weight gain and that their health care providers gave varied recommendations. Factors that were

associated with a total weight gain below the IOM guidelines included lower maternal education, Latina ethnicity, a BMI <19.8 , a lack of provider advice about weight gain, and provider advice to gain weight below the IOM guidelines. Among women with a prepregnancy BMI <19.8 , 51.2% reported a target weight gain below the guidelines versus 10.4% of normal-weight women. Factors independently associated with a target gain above IOM guidelines were a prepregnancy BMI >26 , multiparity, lower age, and provider advice to gain above the IOM guidelines. They concluded that standardization of provider recommendations was necessary and that new approaches were needed to educate providers on gestational weight gain expectations.

The 2 studies still in progress are Peterson *et al.*'s study design³⁷ from 2002 and Althuisen *et al.*'s study³⁸ published in 2006 in the Netherlands. The primary purpose of Peterson *et al.*'s randomized controlled trial is to test the ability of an educational program delivered by community-based professionals to improve diet, activity, and weight loss among new mothers over a 12-month postpartum period and a 6-month maintenance period. Participants were randomized to the Federal Nutrition Program for Women, Infants, and Children or to a multicomponent intervention, including home visits, group classes, and monthly telephone counseling sessions. Women reported decreases in moderate and vigorous physical activity during pregnancy that persisted at 6 months postpartum. Children in the home, longer work hours, and a lack of childcare were predictors of becoming insufficiently active during or after pregnancy. This intervention is still ongoing, with no additional published results to date. Althuisen *et al.* published their design for a randomized control trial using the New Life(style) intervention for pregnant patients. During 5 sessions of individualized counseling using patient education brochures (in weeks 18, 22, 30, and 36 of pregnancy and at 8 weeks postpartum), individual weight gain is discussed with respect to IOM weight gain guidelines. The counseling includes motivation to maintain physical activity and healthy diet practices and notification about personal weight gain. Patients undergoing the intervention will be compared to controls in midwife practices throughout pregnancy and the first postpartum year.

Given current data, more effective interventions seem to be of an individualized approach, but there is a significant lack of data to support any specific model. Thus, more prospective interventional studies are needed. Obese women should be encouraged to begin a regular fitness program preconceptionally and then continue it throughout pregnancy and after delivery. Previously sedentary obese women who

are already pregnant can also be allowed to initiate a moderate program of exercise and physical activity prior to 16 weeks of gestation without compromising fetal growth.³⁹ Validated target heart rate ranges for sedentary overweight and obese pregnant women are 102 to 124 bpm (20–29 years of age) and 101 to 120 bpm (30–39 years of age), as recommended by the American College of Sports Medicine.¹⁰

FUTURE DIRECTIONS

Obesity is an epidemic that affects pregnancy outcomes. It is important to recognize that prepregnancy obesity is more likely to cause obstetric complications than excessive gestational weight gain, which leads to a normal-weight or overweight woman becoming redefined as obese during gestation.¹² On the basis of interventional data that have been reported to date, it appears advantageous to limit weight gain in pregnancy in obese patients to less than 10 lb.⁴¹ As interventions limiting weight gain during pregnancy have been minimal with educational goals at best, further studies are needed to examine intense weight management protocols for obese patients. It has become essential to be proactive about weight management in the prenatal setting, and there is a need for further evaluation of the results of such studies on long-term weight control and other long-term effects on maternal health status.²⁷

At Mount Sinai Hospital, a program for lifestyle modification is currently in use and under investigation in the prenatal clinic setting. Pilot year data support the efficacy of the intervention in limiting weight gain in obese pregnant patients in a small cohort of women. Comparative prospective interventional studies are needed to demonstrate the most efficacious, easily implemented intervention for obese pregnant women. This would help guide the obstetrician/gynecologists in choosing among interventions and in managing this at-risk population.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. *Obesity: Preventing and Managing a Global Epidemic*. Geneva, Switzerland: World Health Organization; 2000: WHO Technical Support Series 894.
2. Ogden CL, Carroll MD, McDowell MA, Flegal KM. *Obesity Among Adults in the United States—No Change Since 2003–2004*. Hyattsville, MD: National Center for Health Statistics; 2007: NCHS Data Brief 1.
3. Flegal KM, Carroll MD, Ogden CL, Johnson CL. Prevalence and trends in obesity among US adults, 1999–2000. *JAMA* 2002; 288: 1723–1727.
4. Hedley AA, Ogden CL, Johnson CL, et al. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *JAMA* 2004; 291: 2847–2850.
5. Siega-Riz AM, Laraia B. The implications of maternal overweight and obesity on the course of pregnancy and birth outcomes. *Matern Child Health J* 2006; 10: S153–S156.
6. Yeh J, Shelton JA. Increasing prepregnancy body mass index: analysis of trends and contributing variables. *Am J Obstet Gynecol* 2005; 193: 1994–1998.
7. Crane JM, White J, Murphy P, et al. The effect of gestational weight gain by body mass index on maternal and neonatal outcomes. *J Obstet Gynaecol Can* 2009; 31: 28–35.
8. Kiel DW, Dodson EA, Artal R, et al. Gestational weight gain and pregnancy outcomes in obese women: how much is enough? *Obstet Gynecol* 2007; 110: 752–758.
9. Weiss JL, Malone FD, Emig D, et al. Obesity, obstetric complications and cesarean delivery rate—a population based screening study. *Am J Obstet Gynecol* 2004; 190: 1091–1097.
10. Mostello D, Kallogjeri D, Tungsiripat R, Leet T. Recurrence of preeclampsia: effects of gestational age at delivery of the first pregnancy, body mass index, paternity, and interval between births. *Am J Obstet Gynecol* 2008; 199: 55.e1–55.e7.
11. Chu SY, Callagan WM, Kim SY, et al. Maternal obesity and risk of gestational diabetes mellitus. *Diabetes Care* 2007; 30: 2070–2076.
12. Sarwer DB, Kelly AC, Gibbons LM, et al. Pregnancy and obesity: a review and agenda for future research. *J Womens Health* 2006; 15: 720–733.
13. Prentice A, Goldberg G. Maternal obesity increases congenital malformations. *Nutr Rev* 1996; 54: 146–150.
14. Cnattingius S, Bergstrom R, Lipworth L, Kramer MS. Prepregnancy weight and the risk of adverse pregnancy outcomes. *N Engl J Med* 1998; 338: 147–152.
15. Rasmussen SA, Chu SY, Kim SY, et al. Maternal obesity and the risk of neural tube defects: a metaanalysis. *Am J Obstet Gynecol* 2008; 198: 611–619.
16. Chu SY, Kim SY, Lau J, et al. Maternal obesity and risk of stillbirth: a metaanalysis. *Am J Obstet Gynecol* 2007; 197: 223–228.
17. Hendler I, Blackwell SC, Bujold E, et al. The impact of maternal obesity on mid trimester sonographic visualization of fetal cardiac and craniospinal structures. *Int J Obstet Relat Metab Disord* 2004; 28: 1607–1611.
18. Paladini D. Sonography in obese and overweight pregnant women: clinical, medicolegal and technical issues. *Ultrasound Obstet Gynecol* 2009; 33: 720–729.
19. Vahratian A, Zhang J, Troendle JF, et al. Maternal prepregnancy overweight and obesity and the pattern of labor progression in term nulliparous women. *Obstet Gynecol* 2004; 104: 943–951.
20. Chu SY, Kim SY, Schmid CH, et al. Maternal obesity and risk of cesarean delivery: a meta-analysis. *Obes Rev* 2007; 8: 385–394.
21. Hibbard JU, Gilbert S, Landon MB, et al. Trial of labor or repeat cesarean delivery in women with morbid

- obesity and previous cesarean delivery. *Obstet Gynecol* 2006; 108: 125–133.
22. Johnson SR, Kolberg BH, Varner MW, Railsback LD. Maternal obesity and pregnancy. *Surg Gynecol Obstet* 1987; 164: 431–437.
 23. Vallejo M. Anesthetic management of the morbidly obese parturient. *Curr Opin Anaesthesiol* 2007; 20: 175–180.
 24. Hood DD, Dewan DM. Anesthetic and obstetric outcome in morbidly obese parturients. *Anesthesiology* 1993; 79: 1210–1218.
 25. Amir LH, Donath S. A systematic review of maternal obesity and breastfeeding intention, initiation and duration. *BMC Pregnancy Childbirth* 2007; 7: 9.
 26. Committee on Nutritional Status During Pregnancy and Lactation. *Nutrition During Pregnancy*. Washington, DC: National Academy Press; 1990: parts I and II.
 27. Weight Gain During Pregnancy: Reexamining the guidelines. IOM Brief Report, May, 2009. Available at: <http://www.iom.edu>. Accessed September 2009.
 28. ACOG Committee Opinion Number 315, September 2005. Available at: <http://www.acog.org>. Accessed September 2009.
 29. Schieve LA, Cogswell ME, Scanlon KS. Maternal weight gain and preterm delivery: differential effects by body mass index. *Epidemiology* 1999; 10: 141–147.
 30. Power ML, Cogswell ME, Schulkin J. US obstetrician-gynaecologist's prevention and management of obesity and pregnancy. *J Obstet Gynaecol* 2009; 29: 373–377.
 31. Lumley J, Oliver SS, Chamberlain C, Oakley L. Interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst Rev* 2004; 4: CD001055.
 32. Smith S, Husley T, Goodnight W. Effects of obesity on pregnancy. *J Obstet Gynecol Neonatal Nurs* 2008; 37: 176–184.
 33. Gray-Donald K, Robinson E, Collier A, et al. Intervening to reduce weight gain in pregnancy and gestational diabetes in Cree communities: an evaluation. *CMAJ* 2000; 163: 1247–1251.
 34. Polley BA, Wing RR, Sims CJ. Randomized control trial to prevent excessive weight gain in pregnancy. *Int J Obes Relat Metab Disord* 2002; 26: 1494–1502.
 35. Olson, C, Strawderman M, Reed RG. Efficacy of an intervention to prevent excessive gestational weight gain. *Am J Obstet Gynecol* 2004; 191: 530–536.
 36. Stotland NE, Haas JS, Brawarsky P, et al. Body mass index, provider advice, and target gestational weight gain. *Obstet Gynecol* 2005; 105: 633–638.
 37. Peterson KE, Sorensen G, Pearson M, et al. Design of an intervention addressing multiple levels of influence on dietary and activity patterns of low-income, postpartum women. *Health Educ Res* 2002; 17: 531–540.
 38. Althuisen E, van Poppel MN, Seidell JC, et al. Design of the New Life(style) study: a randomised controlled trial to optimise maternal weight development during pregnancy. *BMC Public Health* 2006; 6: 168.
 39. Clapp JF III, Kim H, Burciu B, Lopez B. Beginning regular exercise in early pregnancy: effect on fetoplacental growth. *Am J Obstet Gynecol* 2000; 183: 1484–1488.
 40. Davenport MH, Charlesworth S, Vanderspank D, et al. Development and validation of exercise target heart rate zones for overweight and obese pregnant women. *Appl Physiol Nutr Metab* 2008; 33: 984–989.
 41. Asbee SM, Jenkins TR, Butler JR, et al. Preventing excessive weight gain during pregnancy through dietary and lifestyle counseling. *Obstet Gynecol* 2009; 113: 305–311.

A Comprehensive Approach to the Treatment of Uterine Leiomyomata

Karen Stein, MD and Charles Ascher-Walsh, MD

Mount Sinai School of Medicine, New York, NY

OUTLINE

PHARMACOLOGICAL TREATMENT OPTIONS

MINIMALLY INVASIVE TREATMENT OPTIONS

SURGICAL TREATMENT OPTIONS AND COMPARISON STUDIES

CONCLUSION

ABSTRACT

Leiomyomas (fibroids) are the most common tumors in women, with a prevalence between 30% and 50%. They affect women primarily during their reproductive years, spontaneously regressing after menopause in most women. They may cause significant symptoms of pain, dysmenorrhea, abnormal uterine bleeding, and infertility. Because leiomyomas are so common, treatment should be reserved for those patients with symptoms. Treatment options have recently expanded beyond hysterectomy. Medical therapies, including gonadotropin-releasing hormone agonists and progesterone modulators, have become more widely used. Less invasive options such as uterine fibroid embolization, magnetic resonance imaging–guided focused ultrasound, and radiofrequency ablation are being used to avoid more invasive surgery. Because of limited and negative information regarding these alternatives to surgery, they are not recommended for women desiring future fertility. If surgery is desired or required, often less invasive approaches via hysteroscopy for intracavitary lesions or robot-assisted laparoscopy for patients with a small number of myomas have become preferred options. Treatment should be tailored to the

patient. *Mt Sinai J Med* 76:546–556, 2009. © 2009 Mount Sinai School of Medicine

Key Words: myoma, fibroid, treatment, myomectomy, hysterectomy, alternative therapies.

Uterine leiomyomas (commonly called fibroids) are benign muscle cell tumors. They are the most common tumors in women and the most frequent pelvic tumors. The overall prevalence has been estimated at 30% to 50% of women, although one Washington, DC study found uterine fibroids in over 80% of African American women and 70% of white women by the age of 50. Fibroids occur with greatest prevalence during the fifth decade of life and rarely occur before menarche. The tumors are responsive to and dependent on estrogen, with most diminishing in size after menopause. Most fibroids have a poor vascular supply, which may lead to infarction and degeneration. Clinically, about 30% of fibroids become symptomatic, with nulliparous women having a higher likelihood of developing symptoms. The most common symptoms are pressure from an enlarged pelvic mass, pain and dysmenorrhea, and abnormal uterine bleeding. Occasionally, fibroids may cause infertility.¹

Historically, the treatment of fibroids had been limited to surgical options: either hysterectomy or myomectomy. By the 1980s, laparoscopic options became available. In recent years, noninvasive therapies have come to the forefront. These include uterine artery embolization (UAE), magnetic resonance imaging–guided focused ultrasound surgery (MRgFUS), and radiofrequency ablation (RFA). Conservative therapies, too, have become more varied and are readily accessible. Conservative therapies range from contraceptive steroids to progesterone modulators, gonadotropin-releasing hormone (GnRH) agonists, and aromatase inhibitors. These widely divergent therapies, as well as recently published studies regarding their usefulness in the treatment of uterine fibroids, are discussed in this article.

Address Correspondence to:

Charles Ascher-Walsh

Mount Sinai School of Medicine
New York, NY

Email: charles.ascher-walsh@mssm.edu

PHARMACOLOGICAL TREATMENT OPTIONS

Many patients and physicians prefer to treat non-life-threatening conditions with conservative management. Uterine fibroids fall into this category, and subsequently, there are many pharmacological treatments currently used and under study. Contraceptive steroids are often used as first-line therapy for a variety of menstrual abnormalities, including dysmenorrhea and abnormal bleeding in women with and without fibroids. However, there are limited data to support their usage.²

Contraceptive steroids are often used as first-line therapy for a variety of menstrual abnormalities, including dysmenorrhea and abnormal bleeding in women with and without fibroids. However, there are limited data to support their usage.

Combined estrogen/progestin contraceptives may control bleeding and pain symptoms but do not tend to affect fibroid size. Progestin-only therapy has also been studied. In a 2007 in vitro study, Maruo *et al.*³ found that insulin-like growth factor 1 plays a role in the proliferation of leiomyoma cells by up-regulating proliferating cell nuclear antigen and inhibiting apoptosis thorough up-regulation of basal cell lymphoma 2 protein expression. In their experiment, treatment with progesterone significantly decreased IGH-1 messenger RNA expression in cultured leiomyoma cells. However, there were no significant differences in insulin-like growth factor 1 receptor messenger RNA expression between untreated cultures and cultures treated with progesterone. These results may explain progesterone's dual action on leiomyoma cells; one action stimulates their growth, whereas the other inhibits it.

A 2004 study of medroxyprogesterone acetate (Depo Provera) examined a cohort of 20 premenopausal women with menorrhagia attributed to fibroids. All the women received 150 mg per month of medroxyprogesterone acetate. Post-treatment, 30% of the women became amenorrheic, and 70% experienced resolution or improvement of their bleeding patterns. In addition, serum hemoglobin levels were significantly higher post-treatment, with significant decreases in the mean uterine volume and mean

fibroid volumes. This pilot study shows great promise for the use of medroxyprogesterone acetate in providing temporary relief of menorrhagia attributed to fibroids.⁴

Another contraceptive that has proven useful for the treatment of fibroids is the levonorgestrel-releasing intrauterine system (LNG-IUS), which is available in the United States as the Mirena intrauterine device. The LNG-IUS is a small, T-shaped device that is inserted into the uterus and releases about 20 µg of levonorgestrel each day. Its function is to prevent pregnancy by thickening the cervical mucus and preventing the transport of sperm into the endometrial cavity and oviducts.¹ In addition to preventing pregnancy, the LNG-IUS causes amenorrhea in many patients. This beneficial side effect has prompted many researchers to study its usefulness in the treatment of menorrhagia and other symptoms caused by uterine fibroids.

Two studies from 2003 examined the use of the LNG-IUS to treat bleeding related to uterine fibroids. Grigorieva *et al.*⁵ studied 67 premenopausal women with menorrhagia who chose the LNG-IUS as their method of contraception. These authors found that menstrual blood loss, reported as the mean pictorial blood loss assessment score, declined significantly 1 year after insertion. In addition, there were significant increases in ferritin and hemoglobin levels over the year: 18 of 19 who were anemic at the outset were no longer anemic at 12 months. In a similar study of just 19 patients, Mercorio *et al.*⁶ found a gradual decline in the pictorial blood loss assessment score over 12 months, but menorrhagia persisted in 14 patients. These researchers also found a significant decrease in hemoglobin over the 1-year period, in contrast to Grigorieva's group. Several other studies from 2005 to 2007 found results similar to those of Grigorieva *et al.*'s study. Magalhaes *et al.*⁷ studied 3 groups of women: one using the LNG-IUS for the treatment of idiopathic menorrhagia, another using the LNG-IUS for menorrhagia due to leiomyoma, and a control group using the LNG-IUS simply for contraception. This study continued for 3 years and found that uterine volumes decreased in both groups of menorrhagia patients, with a more significant reduction in the fibroid group ($P = 0.041$ for the idiopathic group, $P = 0.014$ for the fibroid group). However, there was no significant change in uterine volume in the control group. After 36 months, amenorrhea and oligomenorrhea were the most common patterns. Amenorrhea was more common in the contraception group and idiopathic menorrhagia group in comparison with the myoma group. Additionally, there were no significant reductions in leiomyoma size over 3 years.

Overall, the studies of LNG-IUS use in women with fibroids have shown mixed results. The majority of studies have shown both a decrease in uterine volume and a significant decrease in bleeding-related symptoms. Therefore, it would be reasonable to use the LNG-IUS in a patient who concurrently desires contraception.

Overall, the studies of LNG-IUS use in women with fibroids have shown mixed results. The majority of studies have shown both a decrease in uterine volume and a significant decrease in bleeding-related symptoms. Therefore, it would be reasonable to use the LNG-IUS in a patient who concurrently desires contraception.

Bayer, Mirena's manufacturer, recommends the use of the device for no longer than 5 years. Thereafter, the device must be replaced.⁸

GnRH agonists present another pharmacological treatment option. These agents lead to amenorrhea in most women and provide a 35% to 65% reduction in leiomyoma volume over 3 months. Currently, the only Food and Drug Administration (FDA)-approved GnRH agonist for the treatment of fibroids is leuprolide as part of a leuprolide plus iron regimen for preoperative treatment in women with large fibroids. Unfortunately, the effects of GnRH agonists are temporary. Post-treatment, myomas gradually regrow to their pretreatment size over a period of several months. Additional concerns with this mode of therapy include the symptoms of pseudomenopause and the adverse effects of hypoestrogenism on bone density. These 2 factors limit the suggested use of GnRH agonists to no more than 6 months, leading to the recommended use of this treatment as an adjuvant to surgical options, especially in anemic women.²

An Italian group studied the long-term effectiveness of the administration of a GnRH agonist plus raloxifene for the treatment of leiomyomas. Raloxifene is a synthetic nonsteroidal drug that is afferent to selective estrogen receptor modulators. It is used to prevent postmenopausal bone loss in women without osteoporosis and to reduce the incidence of fractures in those with osteoporosis. In this study, 50 premenopausal women were treated with

monthly 3.75-mg leuprolide acetate injections and 60 mg of raloxifene daily for 18 cycles. No significant changes in bone mineral density or bone metabolism markers were seen throughout the study. A significant decrease in uterine size (including leiomyoma and nonleiomyoma size) was seen after 6 months, with no further significant changes. This study concluded that the administration of a GnRH agonist plus raloxifene is a safe and effective treatment for uterine leiomyomas.⁹ Raloxifene, however, does not help with other menopausal symptoms such as hot flashes and vaginal dryness, and these persisted throughout the study.

Vercellini and colleagues¹⁰ conducted a controlled trial to examine the impact of GnRH pretreatment on intraoperative blood loss in women undergoing abdominal myomectomy. In this open-label, parallel-group randomized controlled trial, one group of 50 women underwent adjuvant treatment with a GnRH agonist for 2 months, whereas the other group of 50 was scheduled for surgery without this pretreatment. There were no significant differences in intraoperative blood loss or hemoglobin concentrations after the surgery. Jasonni and colleagues¹¹ found a similar result in their study of 66 women who underwent abdominal myomectomy. This study divided participants into 2 groups: 22 patients received long-term leuprolide injections (6 monthly injections prior to surgery), and 16 patients received only 2 monthly injections. Although the uterine volume decreased significantly after 2 injections in both groups, the decrease between 2 and 6 injections was not significant. There were no significant differences in intraoperative blood loss in either group.

Aromatase inhibitors are yet another method of controlling the symptoms caused by uterine fibroids. These agents block ovarian and peripheral estrogen production and decrease estradiol levels after just 1 day of treatment.⁴ In addition, aromatase inhibitors may have fewer side effects than GnRH agonists because of their mechanism of action. As described by Hilario *et al.*,¹² aromatases have the ability to differentially inhibit estrogen action in the leiomyoma while causing less suppression in the ovary. They can therefore reduce the volume of the leiomyoma without the systemic side effects observed with GnRH analogues. This mode of treatment is not yet approved by the FDA but shows great promise. Hilario *et al.* studied a small group (n = 20) of women with large uterus-leiomyoma structures (>300 cc). All the women were treated with anastrozol, a highly selective third-generation nonsteroid substance that does not affect cortisol or aldosterone levels and does not cause androgenic, prostatic, or estrogenic effects. The dose was 1 mg/day for 12 weeks.

Outcome measures were ultrasound examinations at outset, 30 days, and 90 days as well as self-reported assessment of symptoms related to fibroids (menstrual volume, duration of flow, and presence of dysmenorrhea). The mean volume of the uterus-leiomyoma structure was significantly reduced from pretreatment to the 1- and 3-month time points, but there was no significant difference between the 1- and 3-month values. A statistically significant reduction in all self-reported symptoms was found among the 3 periods. Additional findings included no significant changes in the values of follicle-stimulating hormone or estradiol over the 3-month period. This mode of treatment is indeed promising, but larger randomized controlled trials are needed.

Progesterone modulators are perhaps the most widely studied of the pharmacological therapies for symptomatic uterine fibroids. Progesterone receptors are found in high concentrations in leiomyomatous uteri, and antiprogesterone agents work at these receptors. Mifepristone is the most extensively studied of these modulators. Several studies have shown a reduction in leiomyoma volume and a slower rate of recurrent regrowth with high-dose mifepristone administration. Amenorrhea is also a common result of mifepristone use. In contrast to GnRH agonists, mifepristone has the benefit of maintaining bone density. However, this medication carries the risk of simple endometrial hyperplasia and the potential for antiglucocorticoid activity.²

Fiscella and colleagues¹³ conducted a randomized controlled trial of 42 women with significant leiomyoma-related symptoms, a total uterine volume of at least 160 mL, and at least 1 myoma greater than 2.5 cm in size. Twenty-two women received 5 mg of mifepristone once a day for 6 months, whereas the others received placebo. The mifepristone group experienced an improvement in their quality of life by 6 months. By the end of the trial, 41% of the treatment group experienced amenorrhea versus none of the placebo group subjects. Mean hemoglobin levels increased significantly in the treatment group, whereas the subjects in the placebo group experienced a decrease in hemoglobin. Uterine volume decreased an average of 200 mL among women in the treatment group ($P = 0.02$), whereas the placebo group experienced an increase in uterine volume, although this finding was not statistically significant. Adverse effects were not statistically significant between the 2 groups. In 2008, Esteve *et al.*¹⁴ endeavored to determine whether a 5-mg dose of mifepristone provided symptomatic relief and a reduction in uterine volume comparable to those achieved with a 10-mg dose. This study found no significant differences in the reduction of uterine

volume or symptoms between the 2 treatment groups ($n = 100$). It is therefore reasonable to use a 5-mg dose of mifepristone.

Several novel selective progesterone modulators have recently entered the literature. CDB-2914 is one new modulator being studied, and asoprisnil is another. Several small studies of these 2 selective progesterone modulators have shown results similar to those achieved with mifepristone use: reductions in leiomyoma volume, elimination of menstrual bleeding, and improvement in symptomatic quality of life scores among fibroid sufferers.^{15,16} These new modalities show promise but need to be studied in larger groups of patients.

MINIMALLY INVASIVE TREATMENT OPTIONS

In recent years, many nonsurgical treatments for uterine fibroids have been developed and studied. These include UAE (also called uterine fibroid embolization), MRgFUS, and transvaginal radiofrequency thermal ablation. These treatment options are still in the earlier stages of discovery, and no one procedure has been deemed more effective than the others.

UAE is by far the most widely studied of the less invasive methods thus far and has also been proposed as an adjuvant treatment prior to surgery. The technique is usually performed by interventional radiologists. During the procedure, uterine arteries are embolized via a transcutaneous femoral artery approach, and this leads to uterine fibroid devascularization and involution.²

Several large registries from the United States, Canada, and other nations have reported long-term success with this technique.

Several large registries from the United States, Canada, and other nations have reported long-term success [in treating fibroids with UAE].

The Fibroid Registry for Outcomes Data (FIBROID) is one such single-arm, multicenter, longitudinal outcome study for short-term and long-term outcomes of UAE. This registry included 2112 patients at 27 geographically diverse sites, including academic centers, community hospitals, and closed-panel health maintenance organizations. At 36 months, 1196 patients

remained in the study, and 1278 completed the survey. Outcome measures were symptom and health-related scores from the Uterine Fibroid Symptom and Quality of Life Questionnaire. Both symptom scores and quality of life scores were significantly improved at 36 months. In fact, scores for both measures moved into the normal range for the questionnaire. Estimates of hysterectomy, myomectomy, and repeat UAE over the 3 years were as follows: 9.79%, 2.82%, and 1.83% of patients, respectively. The authors concluded that UAE leads to a durable quality of life improvement, with results achievable in any experienced community or academic interventional radiology practice.¹⁷

Another large registry is the Ontario Uterine Fibroid Embolization Trial, which was also a multicenter, prospective, single-arm clinical trial. The Ontario group followed women for 3 months after UAE to determine reductions in fibroid volume, patient-reported symptom improvement, symptom life impact reduction, and treatment satisfaction. The researchers found statistically significant outcomes in all measures, with 91% of patients satisfied with the treatment.^{18,19}

A third large registry asked a slightly different question: the Embolization Versus Hysterectomy (EMMY) trial was a randomized comparison of UAE and hysterectomy. One hundred seventy-seven patients were randomized into 2 groups and followed over a period of 2 years to determine the safety of UAE versus hysterectomy; 23.5% of the UAE patients in this study went on to have hysterectomy. There were no significant differences in improvement versus the baseline in pain and bulk-related complaints. The EMMY trial researchers concluded that UAE has about a 76.5% success rate, which they deemed acceptable.^{20–23} Another 5-year follow-up study of 200 patients treated with UAE found a 20% reoperation rate (hysterectomy, 13.7%; myomectomy, 4.4%; and repeat embolization, 1.6%).^{20–23} These studies, as well as several others that have been conducted, demonstrate that UAE is a safe alternative to hysterectomy for the patient who wishes to retain her uterus.²

These studies, as well as several others that have been conducted, demonstrate that UAE is a safe alternative to hysterectomy for the patient who wishes to retain her uterus.

As previously mentioned, UAE has been proposed as an adjuvant treatment prior to myomectomy. Ustunoz and colleagues²⁴ performed a cross-sectional study of 30 patients with uterine fibroids measuring greater than 10 cm. Fifteen of the patients underwent UAE followed by myomectomy the next day, and the remaining 15 underwent myomectomy only. The UAE group had significantly shorter operating times and significantly lower estimated blood loss. None of the UAE patients needed a transfusion, whereas 2 patients in the myomectomy-only group required a transfusion. The mean hospital stay was significantly shorter in the UAE group. Additionally, 2 patients in the myomectomy-only group required subsequent hysterectomy because of intraoperative complications, and 1 patient suffered a bowel-bladder injury. There were no complications in the UAE group. This study shows UAE's potential for reducing blood loss and minimizing complication rates in women with large fibroids who wish to undergo myomectomy.²⁴

Another burgeoning noninvasive therapy for uterine fibroids is MRgFUS. This modality has been FDA-approved since 2004 and uses high-intensity ultrasound waves directed into a focal volume of a leiomyoma. The ultrasound energy penetrates soft tissue and creates regions of protein denaturation, irreversible cell damage, and coagulative necrosis. Several large studies have examined the short-term outcomes of MRgFUS. Stewart and colleagues²⁵ performed a multicenter clinical trial to determine whether a single treatment session of MRgFUS would reduce fibroid symptoms as measured by the symptom severity score of the Uterine Fibroid Quality-of-Life Instrument. The trial began with 109 premenopausal women with symptomatic uterine leiomyomas and no plans for future pregnancy. At 12 months, 82 women remained in the study. Modest uterine volume reductions were seen at 6 and 12 months (13.5% and 9.5%, respectively); 71% of the women reported symptom reduction at 6 months, and 51% had reduced symptoms at 12 months. The magnitude of symptomatic improvement reported was quite dramatic, with mean decreases of 39% and 36% on symptom severity scores at 6 and 12 months. The incidence of adverse events (heavy menses requiring transfusion, persistent pain and bleeding, nausea, and leg and buttock pain) was low.

Another burgeoning noninvasive therapy for uterine fibroids is MRgFUS. [In a] trial [that] began with 109 premenopausal women, 82 women remained in the study. 71% of the women reported symptom reduction at 6 months, and 51% had reduced symptoms at 12 months. The magnitude of symptomatic improvement reported was quite dramatic, with mean decreases of 39% and 36% on symptom severity scores at 6 and 12 months.

A 2008 study carried out in Japan by Mikami and colleagues²⁶ included 48 patients; this trial examined ablation volumes in the study participants. Successful treatment of the target zone was achieved in 32 of the 48 patients (67%). The remaining 16 patients were either obese or had heterogeneous high-signal intensity in their uterine fibroids on T2-weighted magnetic resonance imaging. The 32 patients who were successfully treated were followed for 6 and 12 months and demonstrated significantly decreased menstrual-related and bulk-related symptoms. This relatively small study shows that MRgFUS is a safe, effective procedure but may not be as useful in obese women or those whose fibroids demonstrate high-signal intensity on T2. Another study of 48 patients found similar outcomes.²⁷ Indeed, this modality shows potential, but more research is needed.

Transvaginal radiofrequency thermal ablation is a new technique that has not yet been approved by the FDA. However, several studies from other nations have demonstrated the therapeutic potential of the procedure. In transvaginal radiofrequency thermal ablation, a radiofrequency needle is inserted into the fibroid. This can be done transvaginally, through the cervical canal, or percutaneously. The tip of the radiofrequency needle is centralized in the myoma and heated to the temperature specified by the particular protocol. The myoma is monitored throughout the ablation, and the duration of ablation is determined by the manufacturer's recommendations or by sonographic criteria.^{28,29} In this technique, the local target tissue is destroyed, but the surrounding tissue is minimally damaged.³⁰

Luo *et al.*³⁰ examined the features of the lesions caused by RFA. Sixty women were treated with RFA through the cervical canal and subsequently underwent hysterectomy. In 30 women, hysterectomy was performed immediately after the treatment, whereas the remaining 30 women received underwent hysterectomy on the third posttreatment day. Leiomyomatous tissue showed coagulative necrosis after RFA treatment. Additionally, interstitial vessels were obliterated, and there was no further expression of estrogen receptor or progesterone receptor.

Recaldini and colleagues³¹ performed a feasibility study examining the safety and efficacy of percutaneous RFA under sonographic guidances. Six premenopausal women with symptomatic uterine fibroids were recruited. All women underwent percutaneous RFA under suprapubic ultrasound guidance. Feasibility was defined as the ability to introduce the needle percutaneously into the target myoma and fully expand the tip to the maximum diameter. Efficacy was determined by improvements in symptoms and quality of life and a reduction of volume of uterine myomas. In all cases, the needle was introduced into the myoma, and the tip was expanded to full size without difficulty. There were no operative or postoperative complications. All women were discharged on the evening of the procedure. Relief of symptoms and reductions in the diameter of the fibroids were measured every 3 months, with a mean follow-up time of 9 months. Fibroids were significantly reduced in size at the last follow-up visit. Four of the 6 patients were free of symptoms at the last follow-up visit, and quality of life scores were significantly improved.

In another Italian feasibility study, 18 women underwent RFA under laparoscopic guidance. The women were evaluated for health-related quality of life scores and size of myomas at 1, 3, 6, 9, and 12 months. The median reductions in fibroid volume were 41.5%, 59%, and 77% at 1, 3, and 6 months, respectively. No further change in fibroid size was observed at 9 and 12 months. Significant improvements in symptom scores and quality of life scores were observed at 3 and 6 months, with 7 of the 9 women who completed 1 year of follow-up becoming completely free of symptoms.³² In a follow-up to this study, 25 patients were followed for 1 year post-RFA, with 18 patients completing 2 years of follow-up and 9 patients completing 3 years of follow-up. The median reduction in myoma volume was 68.8% and 77.9% at 6 months and 1 year, with no further changes in fibroid size at 2 and 3 years. Quality of life measures showed significant and lasting improvement in comparison with the baseline.²⁹

Cho *et al.*²⁸ studied a group of 153 premenopausal women with symptomatic uterine leiomyoma who underwent transvaginal radiofrequency thermal ablation under ultrasound guidance. The reoperation rate was 4.3%. The final reduction rate of the volume of the dominant fibroid was 73%. Symptom scores and health-related quality of life scores showed significant improvement at 18 months post-treatment. Women with myomas greater than 75 cm³ showed higher reoperation rates and lower satisfaction rates.²⁸ These studies show that RFA is, in fact, a feasible and effective technique for the treatment of uterine fibroids. There is still a need for larger studies and randomized trials to examine the true potential of RFA.

SURGICAL TREATMENT OPTIONS AND COMPARISON STUDIES

Hysterectomy is the only definitive treatment for uterine fibroids and has therefore been the most common surgical treatment for many years.² The procedure can be performed via laparotomy, via laparoscopy, or by a vaginal approach. In a woman who desires to keep her uterus, the conservative and nonsurgical options discussed previously may be preferable. Additionally, surgical myomectomy is a procedure that allows a woman to retain her uterus. In a myomectomy, the goal should be to remove all the fibroids while trying to maintain as much uterine integrity as possible. Like hysterectomy, myomectomy can be performed by laparotomy, by laparoscopy, or vaginally.³³

Current trends in medicine are moving toward less invasive treatments options. This move is reflected in the literature, with studies comparing the efficacy of the newer, less invasive treatment modalities for uterine fibroids with hysterectomy and myomectomy. Additionally, several studies have compared myomectomy performed via laparotomy to the same procedure carried out by laparoscopy, whereas others have compared the outcomes in laparoscopic and minilaparotomic myomectomy.

Rossetti *et al.*³⁴ compared laparoscopic and abdominal myomectomy to determine if the former predisposed women to earlier recurrence. Eighty-one patients were randomized to laparoscopic ($n = 41$) or abdominal ($n = 40$) myomectomy. Transvaginal ultrasound was performed prior to surgery, within 15 days of the procedure, and every 6 months for a period of 40 months postoperatively. Another group of 84 patients not participating in the randomized controlled study, but who had also undergone

laparoscopic myomectomy, were also studied for the recurrence rate by the same protocol. Among the randomized patients, there were 9 recurrences in the abdominal group and 11 in the laparoscopic group (23% versus 27%); this difference was not significant. Most of the recurrences were seen within 24 months of surgery. The nonrandomized patients were followed for an average of 36 months. There were 17 recurrences (21.79%), most of which were seen between 10 and 30 months postoperatively. The only significant factor for recurrence was preoperative GnRH agonist treatment ($P < 0.02$). Women pretreated with a GnRH agonist had a higher and delayed recurrence rate in comparison with the untreated patients. This study supports laparoscopic myomectomy when appropriate and corresponds with other studies finding that GnRH may shrink small fibroids to a size not findable during myomectomy and hence predispose women to a quicker recurrence.³⁴ Robotic technology has also enabled surgeons the ability to offer laparoscopic myomectomy to a greater number of patients. One argument against laparoscopic myomectomy is the concern over uterine integrity and uterine rupture in pregnancy. This is due to the fact that with traditional laparoscopic instruments, it is frequently difficult to close the uterine defect after removal of the myoma in the same multilayered fashion that would be used during the same procedure done via laparotomy. Although there is no evidence that supports the idea that a multilayer closure is stronger than a single-layer, bulk closure, many surgeons believe that laparoscopy should mirror the equivalent open procedure as much as possible and feel that traditional laparoscopic myomectomy is not sufficiently equivalent. With the increased dexterity allowed by the robot, the uterine defect can be closed in layers similar to those in laparotomy. For surgeons that believe this to be important, the robot makes laparoscopy possible in some of their patients.

Even though several studies have shown laparoscopic myomectomy to be safe and effective, many surgeons may still favor an open approach for large or multiple fibroids. Minilaparotomy seems to combine some of the advantages of laparoscopy (shorter operating times, minimal blood loss, and short hospitalizations) while still affording the access of laparotomy.³⁵ Palomba and colleagues³⁶ performed a randomized controlled trial to study the surgical outcomes of laparoscopic and minilaparotomic myomectomy. Their study included 136 women at 3 university obstetrics and gynecology departments in Italy. Indications for myomectomy included leiomyoma-related symptoms or unexplained infertility. The results showed similar total operating times

in the 2 groups, although the enucleation time of leiomyomas and suturing time of hysterotomies were statistically longer in the laparoscopic group. Intraoperative blood loss and changes in the hemoglobin level were significantly lower in the laparoscopic group versus the minilaparotomic group; no blood transfusions were necessary. The duration of hospitalization was significantly shorter in the laparoscopic group, but there were no differences in the duration of postoperative ileus or in the time necessary to return to full activity between the 2 groups. The total complication rate was significantly higher in the minilaparotomic group than the laparoscopic group. This study also looked at the locations of myomas, as leiomyoma localization proved to be the main independent variable influencing results. In both groups, the total operative time, time of hysterotomy suturing, and surgical difficulty were significantly higher in patients with posterior or infraligamentary myomas than in patients with myomas in other locations, with no significant differences between the 2 groups. The difficulty of surgery was significantly higher in the laparoscopic group versus the minilaparotomic group when fibroids were located at sites other than the posterior/infraligamentary area. The authors concluded that although laparoscopic myomectomy is a more difficult procedure, it has several advantages over minilaparotomy. Localization of myomas is important in choosing one technique over the other; when fibroids were anterior, fundal, or lateral, minilaparotomic myomectomy was actually simpler and faster than laparoscopy, with no significant differences in intraoperative blood loss, duration of ileus, or time of hospitalization.³⁶

A more recent study by Wen and colleagues³⁷ examined short-term outcomes of leiomyomas treated by myomectomy through conventional laparotomy or ultraminilaparotomy. One hundred thirteen patients were included in this controlled, nonrandomized clinical study; 72 underwent myomectomy by laparotomy, and 41 underwent myomectomy by ultraminilaparotomy. The follow-up period for this study was 1 year. The inclusion criteria and measured outcomes were similar to those of Palomba *et al.*'s study.³⁶ In Palomba *et al.*'s study, the incision in the minilaparotomy group was 4 to 7 cm; in Wen *et al.*'s study, the ultraminilaparotomy incision was 2 to 4 cm, and the incision in the laparotomy group was an 8- to 12-cm Pfannenstiel incision. Wen *et al.* found no significant differences in the operative time, blood loss, maximal postoperative fever, or duration of fever between the 2 groups. Postoperative pain control was via continuous pain-control anesthesia or meperidine. Significantly fewer patients in the ultraminilaparotomy group received pain-control

anesthesia, but significantly more patients in the same group requested meperidine. The time to normal bowel movement was significantly shorter in the ultraminilaparotomy group versus the conventional laparotomy group. Hospital stays were significantly shorter in the ultraminilaparotomy group. Like Palomba *et al.*'s study, there were no significant differences in symptom relief after surgery. Wen's group concluded that the ultraminilaparotomy technique offers some advantages to patients without compromising the therapeutic effects.

As discussed in the noninvasive section, the EMMY trial compared the 2-year efficiency of UAE with hysterectomy in the treatment of menorrhagia caused by uterine fibroids. Although symptomatic palomeasures and health-related quality of life scores were significantly improved in both groups, UAE failed in 23.5% of patients, and this prompted those patients to undergo hysterectomy. The authors concluded that

UAE is a valuable alternative treatment for symptomatic uterine fibroids. Nevertheless, when patients seek for certainty on the cessation of bleeding problems, a hysterectomy remains the treatment of choice.²²

The Randomized Trial of Embolization Versus Surgical Treatment for Fibroids (REST) investigators³⁸ conducted a similar randomized controlled trial comparing the efficacy and safety of UAE and surgical methods. The surgical group included 43 women undergoing hysterectomy and 8 undergoing myomectomy, with 106 patients in the UAE group. There were no significant differences in quality of life scores at 1 year between the groups. Symptom scores were better in the surgical group at 1-year follow-up. However, the embolization group had shorter median hospital stays (1 day versus 5 days) and a shorter time before returning to work. There was no significant difference in adverse events between the 2 groups. Most notably, 21 patients (20%) in the embolization group experienced inadequate symptom control, requiring repeat embolization or hysterectomy.

Another study looked at long-term outcomes of myomectomy and UAE. Broder and colleagues³⁹ surveyed 51 women who underwent UAE and 30 women who underwent abdominal myomectomy at a single institution. The surveys were mailed out 3 or more years post-treatment. Embolization patients were more likely to have pursued further invasive treatment (29% versus 3%). Among women who did not need further surgery, overall symptoms improved in 92% of embolization patients and in 90% of myomectomy patients; this difference was not significant. Goodwin *et al.*⁴⁰ found similar results,

with significant improvements in quality of life scores in both groups. These researchers also reported significantly shorter hospital stays and a shorter time to return to work in the UAE group.

Mara and colleagues,⁴¹ a Czech group, conducted a randomized controlled trial comparing UAE and myomectomy. They studied 121 women with reproductive plans; 58 were randomized to embolization, and 63 were randomized to myomectomy (42 to laparoscopic myomectomy and 21 to open myomectomy; this depended on the size and location of the fibroids). All patients were followed for at least 12 months, although the mean follow-up time was 24.9 months. Just as in the other trials discussed, symptomatic relief was not significantly different between the 2 groups. Although regrowth or recurrence of fibroids occurred at similar rates in the 2 groups, reinterventions were significantly more common in the UAE group, with 19 women requiring secondary myomectomy (32.8% versus 3.2% in the myomectomy group). Fifteen of the cases requiring secondary myomectomy were attributed to the persistence of a large fibroid, with 4 cases attributed to regrowth of a fibroid. At the time this article was written, 26 women from the UAE group and 40 from the myomectomy group had tried to conceive. Thirteen of the UAE patients and 31 of the myomectomy patients had conceived. Four women in the UAE group and 1 in the myomectomy group have become pregnant 2 times; 1 woman after myomectomy gave birth to twins. There were significantly more pregnancies and labors after myomectomy than after UAE. Other obstetric and perinatal results were similar in the 2 groups, although this may be attributed to the low number of pregnancies in the UAE group. The authors concluded that although UAE is as safe and symptomatically effective as myomectomy, myomectomy appears to have superior reproductive outcomes at 2 years post-treatment. The REST investigators³⁸ also reported fertility outcomes, although their results were not as detailed. In their cohort, 8 pregnancies occurred in 5 women (7 in the embolization group and 1 in the myomectomy group). Four pregnancies resulted in miscarriage, 3 resulted in live births, and 1 resulted in the intrauterine death of the fetus at 33 weeks' gestation. Fertility outcomes from Palomba *et al.*'s study⁴² comparing laparoscopic and minilaparotomic myomectomy were also reported. In their study, women were studied for 12 months after the procedure. No differences were found between the 2 groups with respect to cumulative pregnancy and live-birth rates. However, the pregnancy and live-birth rates per cycle were significantly higher in the laparoscopic group than in the minilaparotomic

group. In addition, the time to first pregnancy was significantly lower in the laparoscopic group. In patients with symptomatic leiomyomas, the cumulative pregnancy rate and pregnancy and live-birth rates per cycle were significantly higher, whereas the times to first pregnancy and live birth were significantly shorter, after laparoscopic myomectomy versus minilaparotomic myomectomy. However, in patients with unexplained infertility, there were no significant differences in any of the reproductive outcome assessments between the laparoscopic and minilaparotomic groups.

CONCLUSION

Uterine leiomyomas are the most common pelvic tumors in women, and as such, there are many options for their treatment. Hysterectomy is the only definitive treatment for the condition. However, many patients with myomas are asymptomatic, whereas others wish to retain their uteri for childbearing purposes. As our medical society seeks more conservative measures for the treatment of benign conditions, new, noninvasive therapies are being developed and studied. Treatment options should be tailored to a patient's presentation and her desire for future fertility.

Uterine leiomyomas are the most common pelvic tumors in women, and as such, there are many options for their treatment. Hysterectomy is the only definitive treatment for the condition. However, many patients with myomas are asymptomatic, whereas others wish to retain their uteri for childbearing purposes. As our medical society seeks more conservative measures for the treatment of benign conditions, new, noninvasive therapies are being developed and studied. Treatment options should be tailored to a patient's presentation and her desire for future fertility.

All the conservative and pharmacological therapies discussed here are ultimately temporizing measures, although they may be useful as adjuvant treatments prior to surgery. Of the noninvasive therapies, the only one that has been well studied and shows long-term outcomes is UAE. It is a safe and effective option for women who want to preserve their uteri and attempt to conceive.² MRgFUS and RFA show great promise, but there is a need for larger, long-term randomized controlled studies to further evaluate these modalities. Abdominal myomectomy has been well studied and is considered a safe and effective alternative to hysterectomy for the treatment of symptomatic fibroids.² Recent literature supports the use of laparoscopic myomectomy, and studies have shown comparable results for laparoscopic and minilaparotomic myomectomy.³⁶ There is still a need for further studies in order to establish parameters for choosing an approach for myomectomy.² Finally, fertility outcomes after the various uterus-sparing surgical and nonsurgical approaches should be studied further.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

- Katz VS, Lentz G, Lobo R, Gershenson D. *Comprehensive Gynecology*. 5th ed. Philadelphia, PA: Mosby-Elsevier; 2007.
- American College of Obstetricians and Gynecologists. Alternatives to hysterectomy in the management of leiomyomas. ACOG practice bulletin no. 96. *Obstet Gynecol* 2008; 112: 387–400.
- Maruo T, Ohara N, Matsuo H, et al. Effects of levonorgestrel-releasing IUS and progesterone receptor modulator PRM CDB-2914 on uterine leiomyomas. *Contraception* 2005; 75: S99–S103.
- Venkatachalam S, Bagratee JS, Moodley J. Medical management of uterine fibroids with medroxyprogesterone acetate (Depo Provera): a pilot study. *J Obstet Gynaecol* 2004; 24: 798–800.
- Grigorieva V, Chen-Mok M, Tarasova M, Mikhailov A. Use of a levonorgestrel-releasing intrauterine system to treat bleeding related to uterine leiomyomas. *Fertil Steril* 2003; 79: 1194–1198.
- Mercorio F, De Simone R, Di Spezio Sardo A, et al. The effect of levonorgestrel-releasing intrauterine device in the treatment of myoma-related menorrhagia. *Contraception* 2003; 67: 277–280.
- Magalhaes J, Mendes Aldrighi J, Rodrigues de Lima G. Uterine volume and menstrual patterns in users of the levonorgestrel-releasing intrauterine system with idiopathic menorrhagia or menorrhagia due to leiomyomas. *Contraception* 2007; 75: 193–198.
- Mirena Healthcare Professionals. Placement & removal instructions. Available at: <http://www.mirena-us.com/physician/placement.html>. Accessed September 2009.
- Palomba S, Orio F Jr, Russo T, et al. Long-term effectiveness and safety of GnRH agonist plus raloxifene administration in women with uterine leiomyomas. *Hum Reprod* 2004; 19: 1308–1314.
- Vercellini P, Trespidi L, Zaina B, et al. Gonadotropin-releasing hormone agonist treatment before abdominal myomectomy: a controlled trial. *Fertil Steril* 2003; 79: 1390–1395.
- Jasonni VM, D'Anna R, Mancuso A, et al. Randomized double-blind study evaluating the efficacy on uterine fibroids shrinkage and on intra-operative blood loss of different length of leuprolide acetate depot treatment before myomectomy. *Acta Obstet Gynecol Scand* 2001; 80: 956–958.
- Hilario S, Bozzini N, Borsari R, Baracat E. Action of aromatase inhibitor for treatment of uterine leiomyoma in perimenopausal patients. *Fertil Steril* 2009; 91: 240–243.
- Fiscella K, Eisinger S, Meldrum S, et al. Effect of mifepristone for symptomatic leiomyomata on quality of life and uterine size: a randomized controlled trial. *Obstet Gynecol* 2006; 108: 1381–1387.
- Esteve JLC, Acosta R, Heredia B, et al. Mifepristone for the treatment of uterine leiomyomas. *Obstet Gynecol* 2008; 112: 1029–1036.
- Levens ED, Potlog-Nahari C, Armstrong AY, et al. CDB-2914 for uterine leiomyomata treatment: a randomized controlled trial. *Obstet Gynecol* 2008; 111: 1129–1136.
- Wilkens J, Chwalisz K, Han C, et al. Effects of the selective progesterone receptor modulator asoprisnil on uterine artery blood flow, ovarian activity and clinical symptoms in patients with uterine leiomyomata scheduled for hysterectomy. *J Clin Endocrinol Metab* 2008; 93: 4664–4671.
- Goodwin S, Spies J, Worthington-Kirsch R, et al. Uterine artery embolization for treatment of uterine leiomyomata: long-term outcomes from the FIBROID registry. *Obstet Gynecol* 2008; 111: 22–33.
- Pron G, Cohen M, Soucie J, et al. The Ontario Uterine Fibroid Embolization Trial. Part A. Baseline patient characteristics, fibroid burden, and impact on life. *Fertil Steril* 2003; 79: 112–119.
- Pron G, Bennett J, Common A, et al. The Ontario Uterine Fibroid Embolization Trial. Part 2. Uterine fibroid reduction and symptom relief after uterine artery embolization for fibroids. *Fertil Steril* 2003; 79: 120–127.
- Hehenkamp WJK, Volkers NA, Birnie E, et al. EMMY trial: a randomized comparison of uterine artery embolization and hysterectomy for the treatment of symptomatic uterine fibroids. *Int Congr Ser* 2005; 1279: 179–183.
- Hehenkamp WJK, Volkers NA, Donderwinkel PRJ, et al. Uterine artery embolization versus hysterectomy in the treatment of symptomatic uterine fibroids (EMMY trial): peri- and postprocedural results from a randomized controlled trial. *Am J Obstet Gynecol* 2005; 193: 1618–1629.
- Volkers NA, Hehenkamp WJK, Birnie E, et al. Uterine artery embolization versus hysterectomy in the treatment of symptomatic uterine fibroids: 2 years'

- outcome from the randomized EMMY trial. *Am J Obstet Gynecol* 2007; 196: 519e.1–519.e11.
23. Hehenkamp WJ, Volkers NA, Birnie E, et al. Symptomatic uterine fibroids: treatment with uterine artery embolization or hysterectomy—results from randomized clinical embolization versus hysterectomy (EMMY) trial. *Radiology* 2008; 246: 823–832.
 24. Ustunsoz B, Sahin Ugurel M, Bozlar U, et al. Is uterine artery embolization prior to myomectomy for giant fibroids helpful? *Diagn Interv Radiol* 2007; 13: 210–212.
 25. Stewart EA, Rabinovici J, Tempany CMC, et al. Clinical outcomes of focused ultrasound surgery for the treatment of uterine fibroids. *Fertil Steril* 2006; 85: 22–29.
 26. Mikami K, Murakami T, Okada A, et al. Magnetic resonance imaging-guided focused ultrasound ablations of uterine fibroids: early clinical experience. *Radiat Med* 2008; 26: 198–205.
 27. Morita Y, Ito N, Hikida H, et al. Non-invasive magnetic resonance imaging-guided focused ultrasound treatment for uterine fibroids—early experience. *Eur J Obstet Gynecol Reprod Biol* 2008; 139: 199–203.
 28. Cho HH, Kim J, Kim M. Transvaginal radiofrequency thermal ablation: a day-care approach to symptomatic uterine myomas. *Aust N Z J Obstet Gynecol* 2008; 48: 296–301.
 29. Ghezzi F, Cromi A, Bergamini V, et al. Midterm outcome of radiofrequency thermal ablation for symptomatic uterine myomas. *Surg Endosc* 2007; 21: 2081–2085.
 30. Luo X, Shen Y, Song W, et al. Pathologic evaluation of uterine leiomyoma treated with radiofrequency ablation. *Int J Gynecol Obstet* 2007; 99: 9–13.
 31. Recaldini C, Carrafiello G, Lagana D, et al. Percutaneous sonographically guided radiofrequency ablation of medium-sized fibroids: feasibility study. *AJR Am J Roentgenol* 2007; 189: 1303–1306.
 32. Bergamini V, Ghezzi F, Cromi A, et al. Laparoscopic radiofrequency thermal ablation: a new approach to symptomatic uterine myomas. *Am J Obstet Gynecol* 2005; 192: 768–773.
 33. Plotti G, Plotti F, Di Giovanni A, et al. Feasibility and safety of vaginal myomectomy: a prospective pilot study. *J Minim Invasive Gynecol* 2008; 15: 166–171.
 34. Rossetti A, Sizzi O, Soranna L, et al. Long-term results of laparoscopic myomectomy: recurrence rate in comparison with abdominal myomectomy. *Hum Reprod* 2001; 16: 770–774.
 35. Glasser M. Minilaparotomy myomectomy: a minimally invasive alternative to for the large fibroid uterus. *J Minim Invasive Gynecol* 2005; 12: 275083.
 36. Palomba S, Zupi E, Russo T, et al. A multicenter randomized, controlled study comparing laparoscopic versus minilaparotomic myomectomy: short-term outcomes. *Fertil Steril* 2007; 88: 942–951.
 37. Wen KC, Sung PL, Chao KC, et al. A prospective short-term evaluation of uterine leiomyomas treated by myomectomy through conventional laparotomy or ultraminilaparotomy. *Fertil Steril* 2008; 90: 2361–2366.
 38. Moss JG, Edwards RD, Lumsden MA, et al. Uterine artery embolization versus surgery for symptomatic uterine fibroids: the REST investigators. *N Engl J Med* 2007; 356: 360–370.
 39. Broder MS, Goodwin S, Chen G, et al. Comparison of long-term outcomes of myomectomy and uterine artery embolization. *Obstet Gynecol* 2002; 100: 864–868.
 40. Goodwin SC, Bradley LD, Lipman JC, et al. Uterine artery embolization versus myomectomy: a multicenter comparative study. *Fertil Steril* 2006; 85: 14–21.
 41. Mara M, Maskova J, Fucikova Z, et al. Midterm clinical and first reproductive results of a randomized controlled trial comparing uterine fibroids embolization and myomectomy. *Cardiovasc Intervent Radiol* 2008; 31: 73–85.
 42. Palomba S, Zupi E, Falbo A, et al. A multicenter randomized, controlled study comparing laparoscopic versus minilaparotomic myomectomy: reproductive outcomes. *Fertil Steril* 2007; 88: 933–941.

Emerging Roles of Kruppel-Like Factor 6 and Kruppel-Like Factor 6 Splice Variant 1 in Ovarian Cancer Progression and Treatment

Analisa DiFeo, PhD, Goutham Narla, MD, PhD, and John A. Martignetti, MD, PhD

Mount Sinai School of Medicine, New York, NY

OUTLINE

FRAMEWORK OF GENE DISCOVERY IN EPITHELIAL OVARIAN CANCER

KRUPPEL-LIKE FACTOR 6 AND KRUPPEL-LIKE FACTOR 6 SPLICE VARIANT 1 IN HUMAN CANCERS

KRUPPEL-LIKE FACTOR 6 AND KRUPPEL-LIKE FACTOR 6 SPLICE VARIANT 1 IN EPITHELIAL OVARIAN CANCER

EFFECTS OF KRUPPEL-LIKE FACTOR 6 AND KRUPPEL-LIKE FACTOR 6 SPLICE VARIANT 1 INHIBITION

Clinically Relevant Cancer Cell Phenotypes

E-Cadherin Expression

Vascular Endothelial Growth Factor Expression

FUTURE DIRECTIONS: KRUPPEL-LIKE FACTOR 6 SPLICE VARIANT 1 IS A NOVEL THERAPEUTIC TARGET

ABSTRACT

Epithelial ovarian cancer is one of the most lethal gynecologic cancers and the fifth most frequent cause of female cancer deaths in the United States. Despite dramatic treatment successes in other cancers through the use of molecular agents targeted against genetically defined events driving cancer

development and progression, very few insights into epithelial ovarian cancer have been translated from the laboratory to the clinic. If advances are to be made in the early diagnosis, prevention, and treatment of this disease, it will be critical to characterize the common and private (personalized) genetic defects underlying the development and spread of epithelial ovarian cancer. The tumor suppressor Kruppel-like factor 6 and its alternatively spliced, oncogenic isoform, Kruppel-like factor 6 splice variant 1, are members of the Kruppel-like zinc finger transcription factor family of proteins, which have diverse roles in cellular differentiation, development, proliferation, growth-related signal transduction, and apoptosis. Inactivation of Kruppel-like factor 6 and overexpression of Kruppel-like factor 6 splice variant 1 have been associated with the progression of a number of human cancers and even with patient survival. This article summarizes our recent findings demonstrating that a majority of epithelial ovarian cancer tumors have Kruppel-like factor 6 allelic loss and decreased expression coupled with increased expression of Kruppel-like factor 6 splice variant 1. The targeted reduction of Kruppel-like factor 6 in ovarian cancer cell lines results in marked increases in cell proliferation, invasion, tumor growth, angiogenesis, and intraperitoneal dissemination in vivo. In contrast, the inhibition of Kruppel-like factor 6 splice variant 1 decreases cellular proliferation, invasion, angiogenesis, and tumorigenicity; this provides the rationale for its potential therapeutic application. These results and our recent demonstration that the inhibition of Kruppel-like factor 6 splice variant 1 can dramatically prolong survival in a preclinical mouse model of ovarian cancer are reviewed and discussed. *Mt Sinai J Med* 76:557–566, 2009. © 2009 Mount Sinai School of Medicine

Key Words: Kruppel-like factor 6, Kruppel-like factor 6 splice variant 1, molecular pathogenesis, ovarian cancer.

Address Correspondence to:

John A. Martignetti

Mount Sinai School of Medicine

New York, NY

Email: john.martignetti@mssm.edu

Epithelial ovarian cancer (EOC) is one of the most lethal gynecologic cancers and the fifth most frequent cause of female cancer deaths in the United States.¹ Despite dramatic treatment successes in other cancers through the use of molecular agents targeted against the genetically defined, specific alterations driving cancer development and progression, very few insights into EOC have been translated from the laboratory to the clinic, and overall mortality rates remain dismal. If advances are to be made in the early diagnosis, prevention, and treatment of this disease, it will be critical to characterize the genetic defects underlying the development and spread of ovarian cancer. Such an approach will lead to the identification of biomarkers allowing both earlier detection and the development of personalized therapeutics.

The tumor suppressor Kruppel-like factor 6 (KLF6) and its alternatively spliced, oncogenic isoform, Kruppel-like factor 6 splice variant 1 (KLF6-SV1), are members of the Kruppel-like zinc finger transcription factor family of proteins, which have diverse roles in the regulation of cellular differentiation, development, proliferation, growth-related signal transduction, and apoptosis.^{2,3} KLF6 inactivation and KLF6-SV1 overexpression have been associated not only with the progression of a number of human cancers but more importantly with patient survival.^{4–19} This article summarizes our recent findings demonstrating that a majority of EOC tumors

KLF6 inactivation and KLF6-SV1 overexpression have been associated not only with the progression of a number of human cancers but more importantly with patient survival.

have KLF6 allelic loss and decreased expression coupled with increased expression of KLF6-SV1. The targeted reduction of KLF6 in ovarian cancer cell lines results in marked increases in cell proliferation, invasion, tumor growth, angiogenesis, and intraperitoneal dissemination in vivo. In contrast, KLF6-SV1 inhibition decreases cellular proliferation, invasion, angiogenesis, and tumorigenicity; this highlights its potential therapeutic application. These results are reviewed, and the possibility of therapeutically targeting KLF6-SV1 for the treatment of EOC is discussed.

FRAMEWORK OF GENE DISCOVERY IN EPITHELIAL OVARIAN CANCER

As with many other cancers, altered expression and mutations in proto-oncogenes, tumor suppressor genes, and DNA repair genes are believed to be critical to the initiation, progression, and spread of ovarian cancer. In general, ovarian cancers are known to undergo complex cytogenetic aberrations and are often aneuploid.^{20,21} Although numerous studies have attempted to correlate and systematize these long-known cytogenetic findings, the genes that are the targets of these relatively common alterations have not been identified or proven to be bona fide ovarian cancer genes.

One of the main theoretical arguments for this lack of success in gene identification has centered on the fact that past technologies lacked the necessary resolution to identify these genes. Therefore, a number of more recent studies have turned to high-density single nucleotide polymorphism arrays to identify areas of genomic copy number changes. For example, a recent study using a 500,000 element array, with an average intermarker distance of 2500 bp, identified not only large-scale genomic copy number changes but also several hundred small regions of alterations of less than 500 kb.²² These gains and losses would not have been detected by more traditional low-resolution techniques, including cytogenetics, bacterial artificial chromosome array comparative genomic hybridization, microsatellite mapping, and even the earlier versions of single nucleotide polymorphism-based arrays. Importantly, the regions harbored within the small gains and losses identified in this most recent study contain a number of known cancer genes, including fibroblast growth factor receptor 1 and retinoblastoma 1, and this immediately elevates them to candidate status.²² A parallel genome-wide approach has used gene expression profiling to capture the global view of transcriptional changes in ovarian cancer.^{23,24} These studies have primarily focused on defining gene signatures as predictors of clinical outcome^{25–29} or chemotherapy response.^{30–33} Although this offers great promise for future individualized treatment options, validation of the identified signatures in additional sample sets will be required.

Given the technological, cost, and sample set limitations inherent in these global mechanistic studies, an alternative approach has focused on characterizing the genetic mutations and pathways altered in hereditary forms of ovarian cancer. It is estimated that approximately 10% of ovarian carcinomas occur in women with a familial predisposition to

this cancer.³⁴ Two major syndromes are recognized. The first, hereditary breast-ovarian cancer syndrome (and a possible variant, site-specific ovarian cancer), is associated with mutations in breast cancer gene 1 (BRCA1) and breast cancer gene 2 (BRCA2). Estimated lifetime risk ranges from 16% to 60% in women who are carriers for BRCA1 mutations, whereas BRCA2 mutation carriers have estimated risk ranges of 16% to 27%.^{35–37} The second is hereditary nonpolyposis colorectal cancer (Lynch syndrome II), which is characterized by a predisposition to colorectal, endometrial, and ovarian cancers and is associated with mutations in the genes encoding mismatch repair enzymes.³⁸ Unfortunately, the identification of these genes has not yielded great insight into the genetic basis of sporadic cases of EOC, which represent the overwhelming majority of women. BRCA1 and BRCA2 mutations account for approximately 10% of unselected cases. Mutations in the mismatch repair genes account for approximately 2% of cases. The importance, however, of defining the existence of a hereditary predisposition cannot be overstated because this may allow for gynecologic cancer surveillance for those at risk and lead to the optimal therapeutic strategy for those women in whom EOC develops. A number of investigators have provided suggestive evidence for the benefit of prophylactic oophorectomy in females with germline mutations of these genes.^{39–42} A number of caveats do exist and include findings from a recent study that controlled for the type of chemotherapy used, lead-time biases, and mean age at diagnosis; it demonstrated that BRCA-positive patients have a more favorable clinical outcome than nonhereditary EOC patients.⁴³

A third approach to identifying tumor suppressor genes and oncogenes is the candidate gene approach. The candidate gene approach, which selects individual genes for further analysis on the basis of either their function in relevant biological pathways or their known role in other cancers, has yielded an ever increasing number of genes implicated in sporadic ovarian cancer. The caveat in these studies is that the identified genes have risk-associated genetic effects that are relatively small. Apart from providing this increasing list of candidates for future study, these molecular studies have provided evidence that each ovarian tumor subtype has a distinct set of genetic alterations.⁴⁴ For example, mutations in tumor protein 53, v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog, and v-raf murine sarcoma viral oncogene homolog B1 have been reported in serous tumors, whereas endometrioid cancers have been associated with mutations in cadherin-associated protein beta 1 (β -catenin) and

protein tyrosine phosphatase.⁴⁴ These findings indirectly emphasize the fact that substantial molecular heterogeneity exists between the different cancer subtypes.

Using the framework of candidate gene analysis, we recently investigated the role of KLF6 and the alternatively spliced isoform KLF6-SV1 in EOC. As detailed in this article, our results suggest that the altered expression of these 2 genes is both frequent and critical to EOC development and progression.

KRUPPEL-LIKE FACTOR 6 AND KRUPPEL-LIKE FACTOR 6 SPLICE VARIANT 1 IN HUMAN CANCERS

The tumor suppressor KLF6 is a member of the Kruppel-like zinc finger transcription factor family of proteins, which are involved in regulating differentiation, development, cellular proliferation, growth-related signal transduction, and apoptosis.^{2,3} KLF6 was originally cloned and independently isolated by 3 different groups in different tissues: hepatic mesenchymal stellate cells,⁴⁵ placental cells,⁴⁶ and peripheral blood lymphocytes from a B-cell chronic lymphocytic leukemia patient.⁴⁷ Although its physiological role was unknown, the group at Mount Sinai led by Scott Friedman was the first to demonstrate KLF6's role as a tumor suppressor.⁴ Since then, KLF6 inactivation has been implicated in a number of human cancers, including colorectal,⁶ non-small cell lung,⁷ gastric,⁸ astrocytic glioma,^{9,10} nasopharyngeal,¹¹ and hepatocellular carcinomas.^{12–14}

Beyond its role as a tumor suppressor gene, KLF6 has also established a novel paradigm in cancer genetics. We recently demonstrated in one of the largest multi-institutional association studies of prostate cancer that the presence of a germline single nucleotide polymorphism in the KLF6 gene is associated with an increased lifetime prostate cancer risk by increasing the expression of an alternatively spliced, biologically active KLF6 cytoplasmic isoform, KLF6-SV1.¹⁶ KLF6-SV1 is generated by alternative 5'-splice site selection and, because of an out-of-frame splicing event, retains only the N-terminal activation domain of KLF6 but replaces the terminal DNA-binding zinc fingers with a novel 21 amino acid carboxy domain.¹⁶ Intriguingly, although KLF6-SV1 expression is present in both normal and cancerous tissues, the expression of this isoform is significantly up-regulated in multiple cancers, including prostate cancer, glioblastoma, and lung cancer.^{10,18,19} Moreover, its activity directly opposes KLF6's tumor-suppressive effects on

cell proliferation, colony formation, invasion, and in vivo tumor growth.¹⁸

Expression levels of KLF6-SV1 have now been shown to have a prognostic association with lung¹⁸ and prostate¹⁹ cancers. In prostate cancer, increased expression levels of KLF6-SV1 at the time of prostatectomy were associated with a >4-year survival difference in men. Patients with high levels of KLF6-SV1 expression had a median survival of approximately 30 months versus 80 months in patients with low KLF6-SV1 expression.¹⁹ In lung cancer, increased KLF6-SV1 expression levels were associated with a 6.5-year difference in median survival between patients in the lowest and highest tertiles.¹⁸ Taken together, these results suggest not only that KLF6 loss and KLF6-SV1 overexpression occur in human cancer but also that this dysregulation may be predictive of the stage, long-term outcome, and survival in different cancers.

KRUPPEL-LIKE FACTOR 6 AND
KRUPPEL-LIKE FACTOR 6
SPLICE VARIANT 1 IN
EPITHELIAL OVARIAN CANCER

Given these findings, we first began exploring the role of KLF6 and KLF6-SV1 in EOC by determining the frequency of gene loss and then the expression levels of KLF6 and KLF6-SV1 in human tumors.⁴⁹ A total of 68 paired EOC tumor/normal samples were assayed, and allelic loss was present in 70% of serous carcinomas (32/46), the most common histology. In addition, as expected because of the different genetic backgrounds of the various histological subtypes discussed previously, the loss of heterozygosity pattern was subtype-specific (Table 1). KLF6 loss of heterozygosity was also significantly correlated with the tumor stage ($P < 0.025$) and grade ($P < 0.05$), which are determinants of patient survival. As shown in Figure 1, these allelic losses were associated with markedly decreased expression levels of KLF6. Interestingly, in contrast to a number of other cancer types, no KLF6 DNA sequence mutations were detected, and this suggests that changes in KLF6 expression levels are more important in EOC than somatic mutation.

Given that EOC is believed to arise from ovarian surface epithelial cells, we also compared the level of KLF6 expression in these cells to EOC. KLF6 expression was decreased in most high-grade ovarian cancers. Specifically, KLF6 expression in a set of 10 ovarian surface epithelial cells was compared with its expression across 76 stage III EOC samples by

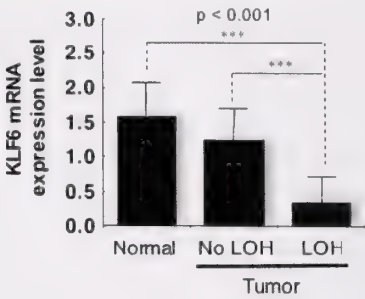


Fig 1. Loss of KLF6 expression in epithelial ovarian cancer tumors with LOH. **Abbreviations:** KLF6, Kruppel-like factor 6; LOH, loss of heterozygosity; mRNA, messenger RNA.

Affymetrix GeneChip analysis. As shown in Figure 2, KLF6 expression was decreased in 62 (82%) tumor samples in comparison with ovarian surface epithelial cells, and the overall decrease was approximately 2-fold ($P < 0.05$). Taken together, these findings are consistent with KLF6's role as an EOC tumor suppressor.

We next explored whether KLF6-SV1 was also expressed in ovarian tissue and whether KLF6 alternative splicing was dysregulated in EOC. RNA from 33 tumors and a panel of 5 normal ovarian tissue samples was analyzed by quantitative reverse-transcript polymerase chain reaction using isoform-specific primers. All tumors except 1 (32/33, 97%) expressed KLF6-SV1. Strikingly, KLF6-SV1 expression was increased nearly 5-fold on average ($P < 0.001$) in the EOC samples (Figure 3, top panel) with respect to normal tissue. In addition, KLF6-SV1 was up-regulated approximately 2-fold in poorly differentiated grade III tumors in comparison with well to moderately differentiated grade I or II tumors ($P < 0.01$; Figure 3, lower panel). Increased KLF6-SV1 expression was particularly notable in view of the corresponding overall decrease in wild-type KLF6 expression.

Table 1. Summary of Allelic Loss of KLF6 in Epithelial Ovarian Cancer.

Histological Type	Allelic Loss of KLF6
Serous	32/46 (70%)*
Endometrioid	4/8 (50%)
Mucinous	1/9 (11%)
Clear cell	0/5 (0%)
Total	37/68 (54%)

Abbreviations: KLF6, Kruppel-like factor 6.
*The allelic loss of KLF6 was significantly more frequently found in serous carcinoma ($P < 0.02$).

EFFECTS OF KRUPPEL-LIKE FACTOR 6 AND KRUPPEL-LIKE FACTOR 6 SPLICE VARIANT 1 INHIBITION

Clinically Relevant Cancer Cell Phenotypes

Because our data suggested that increased KLF6-SV1 expression was present in virtually all serous EOC, we then decided to directly explore the biological role of KLF6-SV1 as it relates to proliferation and invasion/metastasis.⁴⁹ We used small interfering RNA (siRNA), which we previously demonstrated to be highly specific,⁴⁸ to down-regulate KLF6-SV1 expression. Stable ovarian cancer (SKOV-3) cell lines were generated with pSUPER small interfering luciferase (si-luc; a control), pSUPER small interfering Kruppel-like factor 6 (si-KLF6), and pSUPER small interfering splice variant 1 (si-SV1). In these polyclonal pools, KLF6 was reduced approximately 60% and KLF6-SV1 was reduced approximately 75% on protein and RNA levels. Cellular proliferation differences between these siRNA stable cell lines were then measured. In the si-KLF6-expressing cell line, proliferation was increased approximately 2-fold ($P < 0.001$). Conversely, the targeted reduction of KLF6-SV1 resulted in a 25% reduction in proliferation ($P < 0.01$). We next tested the effects of KLF6-SV1 inhibition on invasive capacity. Although targeted KLF6 reduction resulted in a 2.5-fold increase ($P < 0.001$) in cellular invasion, KLF6-SV1 inhibition decreased invasion by approximately 60% ($P < 0.001$; Figure 4).

We next directly examined the role of KLF6 and KLF6-SV1 in tumor growth in vivo. Stable SKOV-3 cell lines expressing si-luc, si-KLF6, or si-SV1 were subcutaneously injected into nude mice, which were monitored weekly for tumor growth and then sacrificed at the end of 6 weeks. Most conspicuously, all mice transplanted with si-SV1 cells ($n = 5$ mice) failed to form persistent tumors. The small tumors that initially developed regressed after 3 weeks. In contrast, reduction of the tumor suppressor KLF6 doubled the tumor growth rate and mass (Figure 4).

The extension of these experiments to define more long-term effects also demonstrated that decreased KLF6, which may occur through a dominant-negative interaction with KLF6-SV1, results in increased intraperitoneal dissemination in vivo. Specifically, the role of KLF6 in EOC progression was examined in an in vivo ovarian cancer mouse model that recapitulates advanced-stage disease with intraperitoneal carcinomatosis and ascites production. The SKOV-3 si-luc or si-KLF6 cell lines were injected into the peritoneal cavity of BALB/c nu/nu

mice, and tumor growth was followed by weekly measurements of abdominal circumferences.

All si-KLF6 mice developed pronounced abdominal swelling from ascites accumulation and intraperitoneal carcinomatosis 6 weeks after tumor cell inoculation, at which point animals from both groups were euthanized, and the tumor burdens were assessed. In the si-KLF6 mice, tumors were present on the peritoneal, intestinal, and diaphragmatic surfaces (Figure 5; arrows). The average ascites volume in the si-KLF6 group was 600 μ L, with an average vascular endothelial growth factor (VEGF) concentration of 870 pg/mL. In marked contrast, no overt metastatic tumors or ascitic fluid collections were observed in the matched si-luc control mice at the same experimental time point.

E-Cadherin Expression

In contrast to well-differentiated ovarian tumors, E-cadherin expression has been shown to be reduced during ovarian tumor progression and metastasis.^{50,51} As shown in Figure 6, the inhibition of KLF6 and the inhibition of KLF6-SV1 have opposite and markedly powerful effects on E-cadherin expression. Although the inhibition of KLF6 decreases E-cadherin by more than half, KLF6-SV1 inhibition increases levels approximately 5-fold. To determine whether E-cadherin is a direct transcriptional target of KLF6, we generated a series of E-cadherin promoter deletion constructs. The sequential deletion of all 4 binding elements within a site having a GGGCGG nucleotide sequence (GC-box) revealed that the -150 and -117 motifs were necessary and sufficient for full KLF6-mediated activation.⁵² To establish a direct interaction between KLF6 and the native E-cadherin promoter in vivo, chromatin immunoprecipitation analysis was also performed. The KLF6 protein occupies the GC-box elements identified in vitro. Therefore, the differences in cellular invasion and proliferation observed in the stable SKOV3 cell lines discussed previously may be due to these KLF6-mediated alterations in E-cadherin.

Vascular Endothelial Growth Factor Expression

The second candidate target that we analyzed was VEGF. Given the striking effects of targeted KLF6 and KLF6-SV1 inhibition on in vivo tumor growth, we examined the effects of their inhibition on the expression of VEGF, a critical regulator of angiogenesis and EOC tumor development.⁴⁹ Differences in all 4 transcripts encoding monomeric VEGF, excluding isoform 121, were examined by quantitative

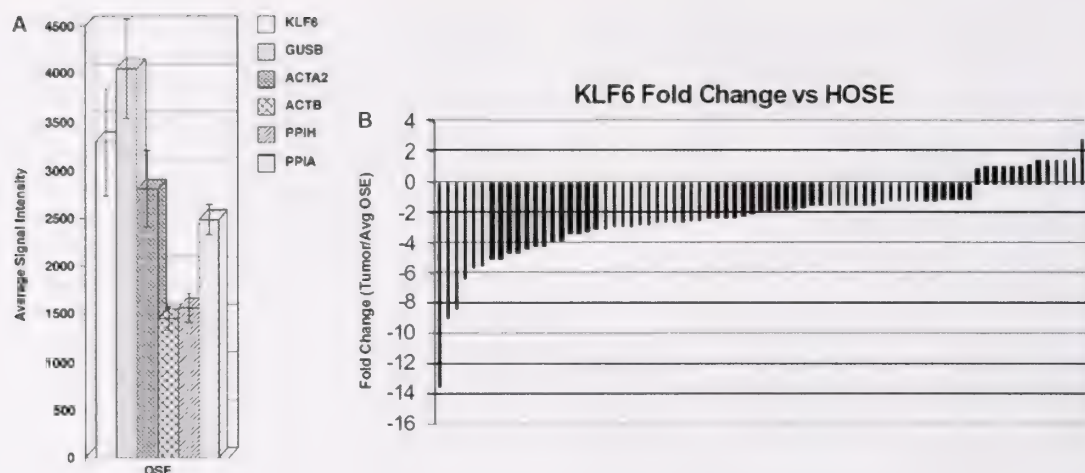


Fig 2. KLF6 expression in OSE brushings versus select housekeeping genes. (A) KLF6 microarray signal intensity values were averaged for 10 OSE brushings and compared to the average expression levels of 5 housekeeping genes: GUSB, ACTA2, ACTB, PPIH, and PPIA. The KLF6 expression level was comparable to the housekeeping gene levels, and this indicates that KLF6 is reliably expressed in ovarian epithelial cells. (B) KLF6 expression in 76 stage III EOC samples versus KLF6 expression in OSE brushings. The KLF6 expression in each microdissected EOC sample was compared to the average expression in all 10 OSE samples. The fold changes were normalized to zero. **Abbreviations:** ACTA2, actin alpha 2 smooth muscle aorta; ACTB, actin beta; EOC, epithelial ovarian cancer; GUSB, glucuronidase beta; HOSE, human ovarian surface epithelial; KLF6, Kruppel-like factor 6; OSE, ovarian surface epithelium; PPIA, peptidylprolyl isomerase A (cyclophilin A); PPIH, peptidylprolyl isomerase H (cyclophilin H).

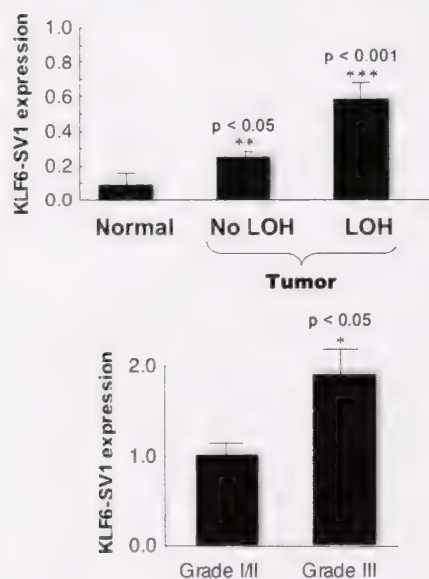


Fig 3. KLF6-SV1 is overexpressed in nearly all epithelial ovarian cancer, regardless of the LOH status, and its expression increases with a higher tumor grade. **Abbreviations:** KLF6-SV1, Kruppel-like factor 6 splice variant 1; LOH, loss of heterozygosity.

reverse-transcript polymerase chain reaction. Inhibition of KLF6-SV1 decreased the secretion of VEGF by approximately 50% (as detected in messenger RNA and protein levels by both Western blotting and enzyme-linked immunosorbent assay) in stable SKOV-3 cell lines in culture. In contrast, KLF6 inhibition led to increased VEGF expression and secretion.

These effects in cultured cells were also sustained in vivo and were associated with tumor growth differences in the mouse xenograft model. There was a significant increase in both VEGF messenger RNA ($P < 0.001$) and protein levels in the si-KLF6-derived tumors. On average, these tumors expressed approximately 15-fold higher VEGF concentrations as measured by enzyme-linked immunosorbent assay.

FUTURE DIRECTIONS: KRUPPEL-LIKE FACTOR 6 SPLICE VARIANT 1 IS A NOVEL THERAPEUTIC TARGET

Ovarian cancer is one of the leading causes of death from gynecologic malignancy, with more than 190,000 new cases each year worldwide. Most patients present with advanced-stage disease, in which the tumor has disseminated/metastasized within the peritoneum, and the cornerstone of treatment is surgical debulking and platinum-based chemotherapy. Although initially responsive to cisplatin, the majority of women will eventually succumb to recurrence and chemoresistance, including 50% of women who have no evidence of disease after primary therapy. Because ovarian cancer recurrences are primarily confined within the peritoneum, from a treatment perspective, the disease can be considered a compartmentalized disease. In fact, the intraperitoneal administration of chemotherapy was

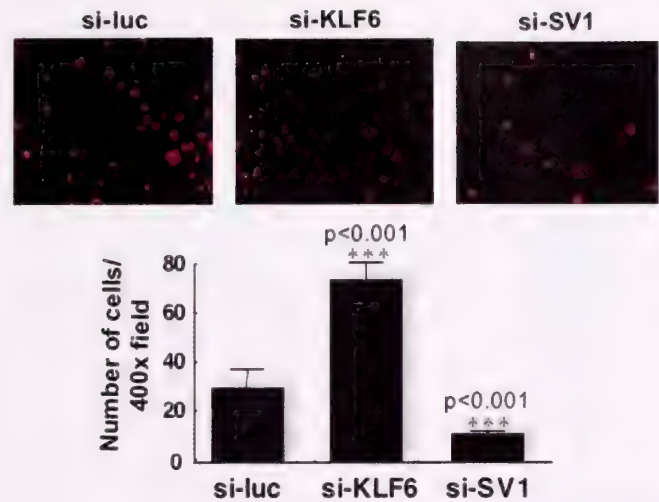


Fig 4. Invasion capacity in small interfering RNA stable cells. The inhibition of KLF6-SV1 decreases tumor invasion and metastasis. The upper panels are representative photomicrographs of the underside of a Matrigel insert stained with 4',6-diamidino-2-phenylindole. The lower panel shows the number of cells per visual field (400× magnification). The cell number was determined from an analysis of 4 representative fields and 3 independent experiments. The statistical analysis was performed with the Student *t* test with a 2-tailed distribution and 2 samples of equal variance. **Abbreviations:** si-KLF6, pSUPER small interfering Kruppel-like factor 6; si-luc, pSUPER small interfering luciferase; si-SV1, pSUPER small interfering splice variant 1.

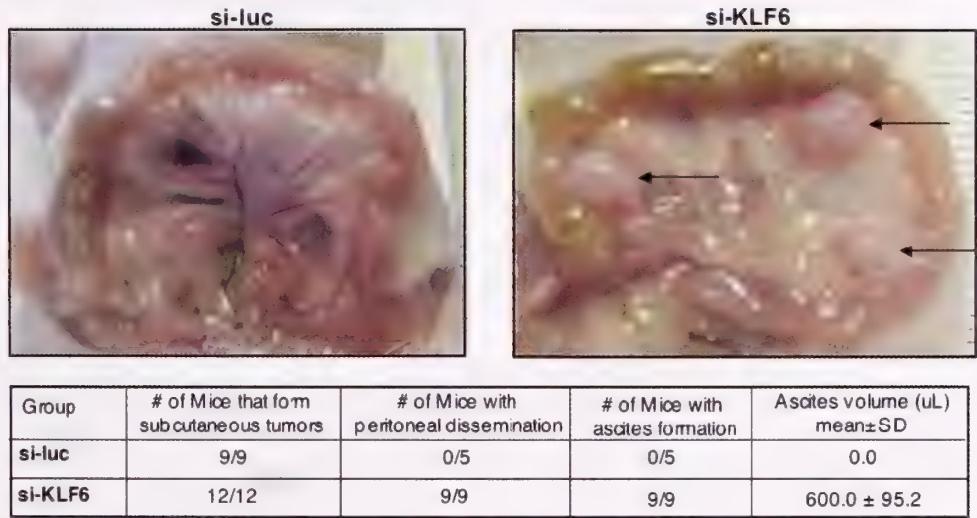


Fig 5. The inhibition of Kruppel-like factor 6 increases intraperitoneal tumor dissemination and vascular endothelial growth factor expression. **Abbreviations:** SD, standard deviation; si-KLF6, pSUPER small interfering Kruppel-like factor 6; si-luc, pSUPER small interfering luciferase.

first proposed 30 years ago,⁵³ and currently used agents, when given via the intraperitoneal route, have been shown to have distinct pharmacokinetic advantages, including higher concentrations and longer half-lives.⁵⁴ Most importantly, numerous studies have suggested that intraperitoneal dosing results in increased patient survival.⁵⁴

Our studies demonstrating overexpression of KLF6-SV1 in EOC and the effect of KLF6-SV1 inhibition on a number of critical tumor phenotypes have led us to propose that KLF6-SV1 represents a novel

Our studies demonstrating overexpression of KLF6-SV1 in

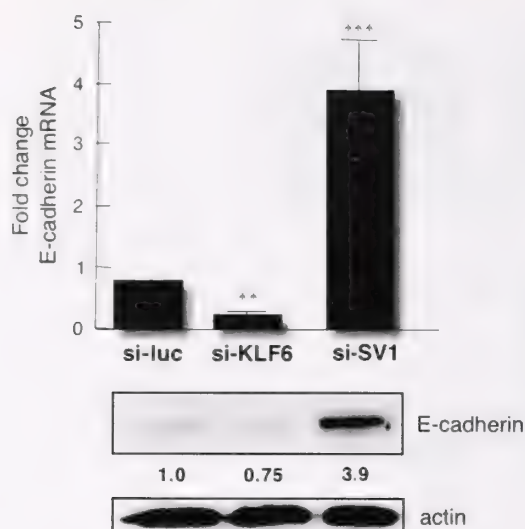


Fig 6. The inhibition of Kruppel-like factor 6 and Kruppel-like factor 6 splice variant 1 has markedly antagonistic effects on E-cadherin expression. **Abbreviations:** mRNA, messenger RNA; si-KLF6, pSUPER small interfering Kruppel-like factor 6; si-luc, pSUPER small interfering luciferase; si-SV1, pSUPER small interfering splice variant 1.

EOC and the effect of KLF6-SV1 inhibition on a number of critical tumor phenotypes have led us to propose that KLF6-SV1 represents a novel therapeutic target in ovarian cancer.

therapeutic target in ovarian cancer. Of particular note with respect to potential therapeutic targeting are our studies demonstrating not only that overexpression of KLF6-SV1 is associated with decreased survival and metastatic spread of certain cancers but also that siRNA-mediated inhibition of KLF6-SV1 has such dramatic effects on tumor behavior in vitro and in vivo.^{19,48,49} Thus, the intraperitoneal delivery of an agent, possibly siRNA-based, to inhibit KLF6-SV1 activity would represent a novel paradigm in targeted ovarian cancer treatment. Because platinum-based chemotherapeutics are a first-line treatment for ovarian cancer, it is particularly notable that KLF6 down-regulation has also been associated with cisplatin resistance in an ovarian cancer cell line.⁵⁵

On the basis of these findings and this rationale, we examined the therapeutic efficacy of an intraperitoneally delivered, next-generation chemically modified siRNA molecule targeting KLF6-SV1.⁵⁶ SKOV3 ovarian cancer cells, which were stably transfected with a luciferase expression vector, were injected into the peritoneum of mice, and the growth and spread of the cells were followed in real time. Survival, the endpoint of the study, was

increased in a dose-dependent manner. Treatment of the mice with the highest dose of si-KLF6-SV1 (20 mg/kg) resulted in a tripling of the median survival from 114 to 311 days, whereas overall survival more than doubled, surpassing 450 days.

On the basis of these findings and this rationale, we examined the therapeutic efficacy of an intraperitoneally delivered, next-generation chemically modified siRNA molecule targeting KLF6-SV1. Treatment of the mice with the highest dose of si-KLF6-SV1 (20 mg/kg) resulted in a tripling of the median survival from 114 to 311 days, whereas overall survival more than doubled, surpassing 450 days.

Mechanistically, KLF6-SV1 was demonstrated to be a prosurvival/antiapoptotic protein that binds to phorbol-12-myristate-13-acetate-induced protein 1, the proapoptotic BH3-only protein, and it targets them both for degradation.⁵⁶

How can these research findings now be translated to the clinic? Given our own work with siRNA, this would seem to be offer a first logical step. Indeed, the horizon for the use of siRNA/RNA interference seems to beckon with the potential for a pharmacopeia of tomorrow for many human diseases.⁵⁷ Already, a number of recent studies have begun

The horizon for the use of siRNA/RNA interference seems to beckon with the potential for a pharmacopeia of tomorrow for many human diseases.

to highlight the potential efficacy and specificity of siRNA/RNA interference-based therapies and have moved well beyond the earliest demonstrations of gene silencing in mammalian cell culture.⁵⁸ For example, a number of investigative groups have demonstrated the in vivo effectiveness of gene silencing in mouse and/or nonhuman primate models of a number of human disorders ranging from respiratory syncytial virus infection⁵⁹ to hypercholesterolemia.⁶⁰ The prospects for the treatment of human cancers have also been long recognized, and delivery issues remain some of the major obstacles.⁶¹ Theoretically,

targeting these disorders by the direct delivery of siRNA into the diseased tissue would represent the most attractive option. The direct delivery route would ensure a high concentration of siRNA and would be predicted to have the least likelihood of unanticipated systemic effects.

Beyond siRNA-mediated treatment, another approach to inhibiting KLF6-SV1 would involve the identification of small molecules that could target it for degradation. The high-throughput screening of large-scale chemical libraries could be used to identify those compounds that selectively lead to KLF6-SV1 inhibition or degradation. On the basis of an ever expanding appreciation of the relevance of KLF6 and KLF6-SV1 to human cancer, the immediate translation of these strategies first to animal models and then to the clinic is a high priority. We believe that developing methods to therapeutically regulate KLF6-SV1 ultimately will play an important role in treating a number of localized and metastatic cancers.

DISCLOSURES

Potential conflict of interest: Drs. Narla and Martignetti are named inventors on patent applications filed by MSSM related to the use of KLF6 and KLF6-SV1 in cancer diagnostics and treatment.

REFERENCES

1. Jemal A, Siegel R, Ward E, et al. *CA Cancer J Clin* 2009; 59: 225–249.
2. Bieker JJ. Kruppel-like factors: three fingers in many pies. *J Biol Chem* 2001; 276: 34355–34358.
3. Black AR, Black JD, Azizkhan-Clifford J. Sp1 and Kruppel-like factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol* 2001; 188: 143–160.
4. Narla G, Heath KE, Reeves HL, et al. KLF6, a candidate tumor suppressor gene mutated in prostate cancer. *Science* 2001; 294: 2563–2566.
5. Chen C, Hytinen ER, Sun X, et al. Deletion, mutation, and loss of expression of KLF6 in human prostate cancer. *Am J Pathol* 2003; 162: 1349–1354.
6. Reeves HL, Narla G, Ogunbiyi O, et al. Kruppel-like factor 6 (KLF6) is a tumor suppressor gene frequently inactivated in colorectal cancer. *Gastroenterology* 2004; 126: 1090–1103.
7. Ito G, Uchiyama M, Kondo M, et al. Kruppel-like factor 6 is frequently down-regulated and induces apoptosis in non-small cell lung cancer cells. *Cancer Res* 2004; 64: 3838–3843.
8. Cho YG, Kim CJ, Park CH, et al. Genetic alterations of the KLF6 gene in gastric cancer. *Oncogene* 2005; 24: 4588–4590.
9. Jeng YM, Hsu HC. KLF6, a putative tumor suppressor gene, is mutated in astrocytic gliomas. *Int J Cancer* 2003; 105: 625–629.
10. Camacho-Vanegas O, Narla G, Teixeira MS, et al. Functional inactivation of the KLF6 tumor suppressor gene by loss of heterozygosity and increased alternative splicing in glioblastoma. *Int J Cancer* 2007; 121: 1390–1395.
11. Chen HK, Liu XQ, Lin J, et al. Mutation analysis of KLF6 gene in human nasopharyngeal carcinomas [in Chinese]. *Ai Zheng* 2002; 21: 1047–1050.
12. Tal-Kremer S, Reeves HL, Narla G, et al. Frequent inactivation of the tumor suppressor Kruppel-like factor 6 (KLF6) in hepatocellular carcinoma. *Hepatology* 2004; 40: 1047–1052.
13. Wang S, Chen X, Zhang W, Qiu F. KLF6 mRNA expression in primary hepatocellular carcinoma. *J Huazhong Univ Sci Technol Med Sci* 2004; 24: 585–587.
14. Kremer-Tal S, Narla G, Chen Y, et al. Downregulation of KLF6 is an early event in hepatocarcinogenesis, and stimulates proliferation while reducing differentiation. *J Hepatol* 2007; 46: 645–654.
15. Peng D, Sheta EA, Powell SM, et al. Alterations in Barrett's-related adenocarcinomas: a proteomic approach. *Int J Cancer* 2008; 1226: 1303–1310.
16. Narla G, DiFeo A, Reeves HL, et al. A germline DNA polymorphism associated with increased prostate cancer risk enhances alternative splicing of the KLF6 tumor suppressor gene. *Cancer Res* 2005; 65: 1213–1222.
17. Teixeira MS, Camacho-Vanegas O, Fernandez Y, et al. KLF6 allelic loss is associated with tumor recurrence and markedly decreased survival in head and neck squamous cell carcinoma. *Int J Cancer* 2007; 121: 1976–1983.
18. DiFeo A, Feld L, Rodriguez E, et al. A functional role for KLF6-SV1 in lung adenocarcinoma prognosis and chemotherapy response. *Cancer Res* 2008; 68: 965–970.
19. Narla G, DiFeo A, Fernandez Y, et al. KLF6-SV1 overexpression accelerates human and mouse prostate cancer progression and metastasis. *J Clin Invest* 2008; 118: 2711–2712.
20. Taetle R, Aickin M, Yang JM, et al. Chromosome abnormalities in ovarian adenocarcinoma: I. Nonrandom chromosome abnormalities from 244 cases. *Genes Chromosomes Cancer* 1999; 25: 290–300.
21. Höglund M, Gisselsson D, Hansen GB, et al. Ovarian carcinoma develops through multiple modes of chromosomal evolution. *Cancer Res* 2003; 63: 3378–3385.
22. Goringe KL, Jacobs S, Thompson ER, et al. High-resolution single nucleotide polymorphism array analysis of epithelial ovarian cancer reveals numerous microdeletions and amplifications. *Clin Cancer Res* 2007; 13: 4731–4739.
23. Konstantinopoulos PA, Spentzos D, Cannistra SA. Gene-expression profiling in epithelial ovarian cancer. *Nat Clin Pract Oncol* 2008; 5: 577–587.
24. Farley J, Ozbun LL, Birrer MJ. Genomic analysis of epithelial ovarian cancer. *Cell Res* 2008; 18: 538–548.
25. Berchuck A, Iversen ES, Lancaster JM, et al. Prediction of optimal versus suboptimal cytoreduction of advanced-stage serous ovarian cancer with the use of microarrays. *Am J Obstet Gynecol* 2004; 190: 910–925.
26. Berchuck A, Iversen ES, Lancaster JM, et al. Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. *Clin Cancer Res* 2005; 11: 3686–3696.
27. Hartmann LC, Lu KH, Linette GP, et al. Gene expression profiles predict early relapse in ovarian cancer after platinum-paclitaxel chemotherapy. *Clin Cancer Res* 2005; 11: 2149–2155.

28. Lancaster JM, Dressman HK, Whitaker RS, et al. Gene expression patterns that characterize advanced stage serous ovarian cancers. *J Soc Gynecol Investig* 2004; 11: 51–59.
29. Spentzos D, Levine DA, Ramoni MF, et al. Gene expression signature with independent prognostic significance in epithelial ovarian cancer. *J Clin Oncol* 2004; 22: 4700–4710.
30. Dressman HK, Berchuck A, Chan G, et al. An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J Clin Oncol* 2007; 25: 517–525.
31. Helleman J, Jansen MP, Span PN, et al. Molecular profiling of platinum resistant ovarian cancer. *Int J Cancer* 2006; 118: 1963–1971.
32. Selvanayagam ZE, Cheung TH, Wei N, et al. Prediction of chemotherapeutic response in ovarian cancer with DNA microarray expression profiling. *Cancer Genet Cytogenet* 2004; 154: 63–66.
33. Spentzos D, Levine DA, Kolia S, et al. Unique gene expression profile based on pathologic response in epithelial ovarian cancer. *J Clin Oncol* 2005; 23: 7911–7918.
34. Brown GJ, St John DJ, Macrae FA, Aittomäki K. Cancer risk in young women at risk of hereditary nonpolyposis colorectal cancer: implications for gynecologic surveillance. *Gynecol Oncol* 2001; 80: 346–349.
35. Brose MS, Rebbeck TR, Calzone KA, et al. Cancer risk estimates for BRCA1 mutation carriers identified in a risk evaluation program. *J Natl Cancer Inst* 2002; 94: 1365–1372.
36. Ford D, Easton DF, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* 1998; 62: 676–689.
37. Bosze P, Bast RC, Berchuck A, et al. Consensus statements on prognostic factors in epithelial ovarian carcinoma. Report of the consensus meeting organized by the European Society of Gynaecological Oncology. *Eur J Gynaecol Oncol* 2000; 21: 513–526.
38. Chung DC, Rustgi AK. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. *Ann Intern Med* 2003; 138: 560–570.
39. Rebbeck TR, Lynch HT, Neuhausen SL, et al., for the Prevention and Observation of Surgical End Points Study Group. Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. *N Engl J Med* 2002; 346: 1616–1622.
40. Kauff ND, Satagopan JM, Robson ME, et al. Risk-reducing salpingo-oophorectomy in women with a BRCA1 or BRCA2 mutation. *N Engl J Med* 2002; 346: 1609–1615.
41. Schmeler KM, Lynch HT, Chen LM, et al. Prophylactic surgery to reduce the risk of gynecologic cancers in the Lynch syndrome. *N Engl J Med* 2006; 354: 261–269.
42. Finch A, Beiner M, Lubinski J, et al. Salpingo-oophorectomy and the risk of ovarian, fallopian tube, and peritoneal cancers in women with a BRCA1 or BRCA2 mutation. *JAMA* 2006; 296: 185–192.
43. Tan DS, Rothermundt C, Thomas K, et al. "BRCAness" syndrome in ovarian cancer: a case-control study describing the clinical features and outcome of patients with epithelial ovarian cancer associated with BRCA1 and BRCA2 mutations. *J Clin Oncol* 2008; 26: 5530–5536.
44. Rabban JT, Bell DA. Current issues in the pathology of ovarian cancer. *J Reprod Med* 2005; 50: 467–467.
45. Kim Y, Ratzu V, Choi SG, et al. Transcriptional activation of transforming growth factor beta1 and its receptors by the Kruppel-like factor Zf9/core promoter-binding protein and Sp1. Potential mechanisms for autocrine fibrogenesis in response to injury. *J Biol Chem* 1998; 273: 33750–33758.
46. Koritschoner NP, Bocco JL, Panzetta-Dutari GM, et al. A novel human zinc finger protein that interacts with the core promoter element of a TATA box-less gene. *J Biol Chem* 1997; 272: 9573–9580.
47. El Roubi S, Rao PH, Newcomb EW. Assignment of the human B-cell-derived (BCD1) proto-oncogene to 10p14-p15. *Genomics* 1997; 43: 395–397.
48. Narla G, DiFeo A, Yao S, et al. Targeted inhibition of the KLF6 splice variant, KLF6 SV1, suppresses prostate cancer cell growth and spread. *Cancer Res* 2005; 65: 5761–5768.
49. DiFeo A, Narla G, Hirshfeld J, et al. Roles of KLF6 and KLF6-SV1 in ovarian cancer progression and intraperitoneal dissemination. *Clin Cancer Res* 2006; 12: 3730–3739.
50. Darai E, Scoazec JY, Walker-Combrouze F, et al. Expression of cadherins in benign, borderline, and malignant ovarian epithelial tumors: a clinicopathologic study of 60 cases. *Hum Pathol* 1997; 28: 922–928.
51. Hudson LG, Zeineldin R, Stack MS. Phenotypic plasticity of neoplastic ovarian epithelium: unique cadherin profiles in tumor progression. *Clin Exp Metastasis* 2008; 25: 643–655.
52. DiFeo A, Narla G, Camacho-Vanegas O, et al. E-cadherin is a novel transcriptional target of the KLF6 tumor suppressor. *Oncogene* 2006; 25: 6026–6031.
53. Dedrick R, Myers C, Bungay PM, et al. Pharmacokinetic rationale for peritoneal drug administration in the treatment of ovarian cancer. *Cancer Treat Rep* 1978; 62: 1–11.
54. Trimble EL, Thompson S, Christian MC, Minasian L. Intraperitoneal chemotherapy for women with epithelial ovarian cancer. *Oncologist* 2008; 13: 403–409.
55. Macleod K, Mullen P, Sewell J, et al. Altered ErbB receptor signaling and gene expression in cisplatin-resistant ovarian cancer. *Cancer Res* 2005; 65: 6789–6800.
56. DiFeo A, Huang F, Sangodkar J, et al. KLF6-SV1 is a novel antiapoptotic protein that targets the BH3-only protein NOXA for degradation and whose inhibition extends survival in an ovarian cancer model. *Cancer Res* 2009; 69: 4733–4741.
57. Gewirtz AM. On future's doorstep: RNA interference and the pharmacopeia of tomorrow. *J Clin Invest* 2007; 117: 3612–3614.
58. Elbashir SM, Harborth J, Lendeckel W, et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 2001; 411: 494–498.
59. Bitko V, Musiyenko A, Shulyayeva O, Barik S. Inhibition of respiratory viruses by nasally administered siRNA. *Nat Med* 2005; 11: 50–55.
60. Frank-Kamenetsky M, Grefhorst A, Anderson NN, et al. Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proc Natl Acad Sci U S A* 2008; 105: 11915–11920.
61. Pai SI, Lin YY, Macaes B, et al. Prospects of RNA interference therapy for cancer. *Gene Ther* 2006; 13: 464–477.

Recent Advances in the Surgical Management of Cervical Cancer

Konstantin Zakashansky, MD,¹ William H. Bradley, MD,^{1,2} Linus Chuang, MD,¹
Jamal Rahaman, MD,¹ and Peter Dottino, MD¹

¹Division of Gynecologic Oncology, Department of Obstetrics, Gynecology, and Reproductive Science,
Mount Sinai School of Medicine, New York, NY

²Department of Obstetrics and Gynecology, Medical College of Wisconsin, Milwaukee, WI

OUTLINE

ABDOMINAL RADICAL HYSTERECTOMY

VAGINAL RADICAL HYSTERECTOMY

LAPAROSCOPIC ASSISTED RADICAL HYSTERECTOMY
AND TOTAL LAPAROSCOPIC RADICAL
HYSTERECTOMY

ROBOTIC LAPAROSCOPIC RADICAL HYSTERECTOMY
FERTILITY-SPARING SURGERY

CONCLUSION

ABSTRACT

Recent advances in the surgical management of early cervical cancer, including abdominal, laparoscopic, vaginal, and robotic approaches to radical hysterectomy as well as fertility-sparing radical trachelectomy, are reviewed. The nerve-sparing abdominal radical hysterectomy technique allows for a significant reduction in postoperative bladder morbidity. Radical vaginal hysterectomy with laparoscopic lymph node dissection is a well-recognized technique that offers excellent cure rates without abdominal entry as well as reduced postoperative febrile and gastrointestinal morbidity. Total laparoscopic radical hysterectomy is a minimally invasive alternative to the

traditional abdominal radical hysterectomy approach and yields a comparable safety profile with a significant reduction in blood loss and hospital stay. Robotic surgery is becoming more widely accepted in the management of gynecologic cancers, including radical hysterectomy for early cervical cancer. Young women desiring to bear children in the future may be candidates for fertility preservation options, and the radical trachelectomy operation has been described and performed with abdominal, vaginal, laparoscopic, and robotic techniques. There are a number of surgical options for the treatment of women with early cervical cancer. The feasibility and safety of some of these techniques have been well established, whereas for others, the oncological outcome data are only preliminary. The decision to use newer techniques should be directed by patient variables as well as the surgeon's training and competence with laparoscopy, robotics, or vaginal surgery. *Mt Sinai J Med* 76:567–576, 2009. © 2009 Mount Sinai School of Medicine

Key Words: cervical cancer, laparoscopy, radical hysterectomy, robotic surgery.

In 2009 11,270 new cases of invasive cervical cancer and 4,070 cancer-related deaths were expected.¹ Radical hysterectomy is the recommended surgical treatment for stage IA1 carcinoma of the cervix with lymph-vascular space involvement as well as stage IA2, IB, and IIA carcinoma.

Radical hysterectomy is the procedure that is most identified with gynecologic oncology as a surgical subspecialty. Its introduction in the management of early cervical cancer at the turn of the century by Schauta² was an important milestone. It allowed for radical extirpation of the parametria with a vaginal approach. Wertheim's development of the abdominal radical technique, however, led to the near abandonment of the vaginal radical hysterectomy,

Address Correspondence to:

Konstantin Zakashansky

Division of Gynecologic
Oncology

Department of Obstetrics
Gynecology, and Reproductive
Science Mount Sinai Medical
Center New York, NY

Email: konstantin.zakashansky
@mssm.edu

despite the 40% mortality rate that was associated with the abdominal procedure.³

With the introduction of radiation as a therapeutic modality for cervical cancer, surgical treatment of this disease declined. In the 1940s, Meigs⁴ reintroduced the Wertheim procedure, combining it with a complete pelvic lymph node dissection to increase its therapeutic efficacy. Antibiotics and improvements in operating techniques and practices reduced the previously described morbidity and mortality of this procedure to acceptable levels. Therefore, the Meigs procedure became the standard of care in the United States for the treatment of cervical cancer. Mitra⁵ used an extraperitoneal pelvic lymphadenectomy technique in combination with radical vaginal hysterectomy in the 1950s. However, it was not until Dargent *et al.*'s introduction of laparoscopic pelvic lymphadenectomy⁶ that radical vaginal hysterectomy was looked at with renewed interest. Dargent *et al.* delineated the benefits of the vaginal approach, which included fewer intraoperative complications, shorter operative times, and a more rapid postoperative recovery.

Advances in laparoscopic instrumentation have made it possible to safely perform radical hysterectomy laparoscopically. Since the original reports of the early 1990s,⁷⁻⁹ total laparoscopic radical hysterectomy with pelvic and/or para-aortic lymph node dissection has been gaining acceptance as a feasible alternative to an open radical procedure.

Since the original reports of the early 1990s, total laparoscopic radical hysterectomy with pelvic and/or para-aortic lymph node dissection has been gaining acceptance as a feasible alternative to an open radical procedure.

Multiple publications have now established both its safety and feasibility.¹⁰ The advantages of laparoscopic radical hysterectomy echo those seen with laparoscopy in general: shorter hospitalization, faster bowel function recovery, less postoperative pain, and decreased overall cost.

Despite the improvements in quality of life outcomes and complication profiles, conventional laparoscopy is not without technical drawbacks; including the use of nonarticulated instruments, inefficient handle-to-tip force transmission, an occasionally awkward and uncomfortable position for

the surgeon at the operating table, and a flat, 2-dimensional image.^{11,12}

Robotic assisted procedures have been evolving rapidly and are now widely used in various surgical fields. The advantages offered by this new technology

Robotic assisted procedures have been evolving rapidly and are now widely used in various surgical fields.

include a 3-dimensional magnified field, tremor filtration, and 5 or 6 degrees of instrument mobility inside the body, thus significantly reducing the ergonomic problems associated with conventional laparoscopic equipment.

The increased use of the cervical cancer screening test has led to the diagnosis of cervical cancer at a younger age. According to the US Census Bureau, 28 of 1000 women and 11 of 1000 women will give birth to their first child between the ages of 30 and 34 and between the ages of 35 and 39, respectively.¹³ As more women are delaying childbearing, fertility preservation has become an important consideration.

In order to avoid performing radical hysterectomy for these women, radical trachelectomy or resection of only the cervix and parametrial tissues was developed. In 1987, French gynecologist Daniel Dargent was the first to develop radical vaginal trachelectomy.¹⁴ Women who desire to preserve fertility with a tumor size of less than 2 cm may be candidates for this procedure.

We will highlight the recent advances in radical hysterectomy as well as radical trachelectomy through the review of relevant literature published recently.

ABDOMINAL RADICAL HYSTERECTOMY

Abdominal radical hysterectomy has undergone little modification since it was originally described by Wertheim³ and Meigs.⁴ It is well established that the morbidity of this procedure is mainly related to the radicality of the parametrial and vaginal resection, which results in pelvic organ denervation. Bladder dysfunction has long been acknowledged as one of the major potential complications of radical (class III) hysterectomy, with a prevalence as high as 80%.¹⁵ Axelsen and Petersen¹⁶ found that symptoms of urinary incontinence or retention persisted up to

10 years in two-thirds of patients who underwent radical hysterectomy without radiation.

Originally developed by Yabuki *et al.*¹⁷ in Japan, nerve-sparing radical hysterectomy has been proposed as a technique that improves quality of life by reducing bladder morbidity. There is no clear

Nerve-sparing radical hysterectomy has been proposed as a technique that improves quality of life by reducing bladder morbidity.

consensus about which parts of the uterine ligaments can be spared during this procedure because many questions still remain regarding pelvic neuroanatomy. Original nerve-sparing techniques concentrated on parasympathetic nerve tissue preservation within the cardinal ligament; today, more and more attention is being paid to the uterosacral ligament with its rich nerve content.¹⁸ Most of the recent techniques aim to preserve the lateral branches by mobilizing the neural tissue of the medial leaf of the broad ligament, allowing the resection of the uterosacral ligament without complete destruction of the hypogastric plexus. Other modifications include preservation of the autonomic nerve supply at the lateral parametrium and anteriorly within the vesicouterine ligament.¹⁹ When using this surgical approach, Charoenkwan *et al.*²⁰ reported 100% resolution of urinary retention (as defined by a postvoid residual of <100 cc of urine) by 28 days versus a 66% rate among historical controls at their institution.

Existing techniques of nerve-sparing radical hysterectomy vary in the extent of the parametrial resection.²¹ Some authors insist that despite modifications allowing for nerve preservation, the radicality of the traditional approach can and must be preserved.²² Others argue that radical parametrectomy has little impact on survival,²³ so nerve sparing can be achieved by a reduction of the radicality of resection (modified radical hysterectomy) with the same oncological efficacy and significantly reduced morbidity. Further studies are required to better define the true clinical value of nerve-sparing radical hysterectomy and its impact on oncological outcome.

VAGINAL RADICAL HYSTERECTOMY

Vaginal radical hysterectomy, pioneered by Schauta at the turn of the 20th century, provided a safe operative approach to the management of

cervical cancer with previously unseen survival rates. The most important disadvantage, however, of the Schauta procedure was the inability to perform lymphadenectomy, which necessitated a second operation. Advancements in laparoscopy have allowed the safe and relatively easy removal of the lymph nodes endoscopically. Furthermore, the improvement of laparoscopic equipment and techniques has expanded the role of laparoscopy to include uterine artery ligation, ureteral tunneling, and cardinal and uterosacral ligament transection. These steps were traditionally performed transvaginally by Schauta and Schuckardt. The classic Schauta procedure today is performed only in a few select centers. Dottino *et al.*²⁴ compared the potential benefits of classical radical vaginal hysterectomy with those of radical abdominal hysterectomy for the treatment of early stage cervical cancer. In this retrospective comparative analysis, 40 patients underwent laparoscopic lymphadenectomy followed by radical vaginal hysterectomy. Vaginal cases were matched 1:1 with patients that underwent abdominal radical hysterectomy with pelvic lymphadenectomy on the basis of age, histological types, Federation Internationale de Gynecologie et d'Obstetrique stages, and tumor volumes. They concluded that the vaginal approach offers faster recovery, shorter operative times, and decreased blood loss. There was 1 intraoperative complication in the vaginal hysterectomy group: a cystotomy that was recognized and repaired with no sequelae. No recurrences were observed at 70 months in the vaginal group, but there were 2 recurrences in the abdominal group. The authors concluded that vaginal radical hysterectomy can provide a satisfactory specimen for therapeutic and prognostic benefits and has a favorable complication profile.

Vaginal radical hysterectomy in combination with laparoscopic lymphadenectomy is a safe approach to the treatment of early stage cervical cancer. Familiarity with the radical vaginal hysterectomy

Vaginal radical hysterectomy in combination with laparoscopic lymphadenectomy is a safe approach to the treatment of early stage cervical cancer.

technique would facilitate performing fertility-sparing radical vaginal trachelectomy in young women with early cervical cancer. Training programs in gynecologic oncology should continue to cultivate

this technique and offer this approach as a surgical option in the management of cervical cancer.

LAPAROSCOPIC ASSISTED RADICAL HYSTERECTOMY AND TOTAL LAPAROSCOPIC RADICAL HYSTERECTOMY

Techniques for performing laparoscopic radical hysterectomy with pelvic and para-aortic lymph node dissection have been described for over 15 years. Although laparoscopy has been widely accepted in the management of corpus cancer, fewer practitioners have adopted a laparoscopic approach to the type III hysterectomy.

To date, there have been over 1000 laparoscopic radical hysterectomies reported in the literature.^{7,10,25-43} When comparing the laparoscopic approach to the abdominal approach, perioperative studies have been remarkably consistent in describing the advantages and disadvantages. Laparoscopic series uniformly have lower blood loss, longer operating times, and either shorter hospital stays or quicker return of bowel function. The safety and feasibility

Laparoscopic series uniformly have lower blood loss, longer operating times, and either shorter hospital stays or quicker return of bowel function.

of the operation have clearly been established. Importantly, an Italian series of 50 laparoscopic radical hysterectomies demonstrated equivalency in specimen size and parametrial width.²⁸

In all comparative series, laparoscopic cases are reported to take longer than open cases. Most commentaries attribute this to a learning curve and report that increased experience will result in a decrease in the time required to perform the operation. After completing the first 26 cases, Spirtos *et al.*⁴¹ reported a reduction in the average operative time from 255 to 186 minutes. The group from Clermont-Ferrand reported their operative time to not be influenced by a body mass index >24 or prior surgery, but they did appreciate a decrease in the operating room time from 290 minutes for the first 25 patients to 226 for the final 25 patients.⁴⁰ Other series have not demonstrated a similar decrease in the operative time, although these series either had lower average operative times or involved fellows in a training program.^{10,38,44} In a fellowship training

setting, Zakashansky *et al.*¹⁰ observed an increase in the median operative time in laparoscopic radical hysterectomies in comparison with open procedures. However, the median operative time of 315 minutes was consistent with that of other academic centers (see Table 1).

Intraoperative complications of the urinary tract tend to be higher with the laparoscopic radical hysterectomy approach versus laparotomy. Steed *et al.*⁴² reported a 10% rate of cystotomies, all of which were repaired without conversion to laparotomy. In the largest series, which included 313 patients, Xu *et al.*⁴³ reported a cystotomy rate of 1.6% with a rate of either vesicovaginal or ureterovaginal fistulas at 2.9%. Such complications may be reflective of the energy source used during the ureteric unroofing and bladder dissection. Unlike open procedures in which blunt dissection with suture ligation is commonly used, laparoscopic procedures require dissecting sources that can result in thermal injuries, leading to either intraoperative or delayed complications.

One important distinction between outcomes in laparoscopic and open procedures appears to be in the return of bladder function. Although the majority of patients regain bladder function at an appropriate interval, patients undergoing laparoscopic hysterectomies tend to have longer bladder atony. Steed *et al.*⁴² reported a 10-day interval

Although the majority of patients regain bladder function at an appropriate interval, patients undergoing laparoscopic hysterectomies tend to have longer bladder atony.

for the return of bladder function for laparoscopic procedures versus 5 days for open procedures. The Gateshead group also had a doubling of the time for return of bladder function from 3 to 7 days.³⁰ Other groups, such as MD Anderson, as reported by Frumavitz *et al.*,²⁷ did not detect a difference; however, the median was prolonged at 13 days for both groups, and this may simply represent a difference in practice style regarding catheter management. More recently, Uccella *et al.*²⁸ provided a report demonstrating no such differences in urinary intraoperative or postoperative complications, indicating that such differences may be reduced as the surgeon's experience increases.

For any oncologist, the ultimate determination of adequacy of the laparoscopic approach is the recurrence rate. Are the removed specimens

Table 1. Reports of Laparoscopic Radical Hysterectomy Versus Abdominal (Open) Radical Hysterectomy from 2006 to 2007

	Frumavitz <i>et al.</i> ²⁷			Ghezzi <i>et al.</i> ²⁸			Morgan <i>et al.</i> ¹¹			Li <i>et al.</i> ³³			Zakashansky <i>et al.</i> ¹⁰		
	Open	Laparo- scopic	Significant Difference	Open	Laparo- scopic	Significant Difference	Open	Laparo- scopic	Significant Difference	Open	Laparo- scopic	Significant Difference	Open	Laparo- scopic	Significant Difference
Number (n)	54	35	No	48	50	No	30	32	No	35	90	No	30	30	No
Operating room time (minutes)	307	344	Yes	260	285	No	131	187	Yes	217	263	Yes	243	319	Yes
Estimated blood loss (cc)	548	319	Yes	500	200	Yes	301	203*	Yes	455	370	No	520	200	Yes
Length of hospital stay	2	5	Yes	10	6	Yes	7.8	5.9	Yes	13.8	13.7	No	5.6	3.8	Yes
Pelvic lymph node	19	14	Yes	23	21	No	17.4	14.8	No	21	19	No	21.8	31	Yes
Recurrence rate (%)	37	2.9	No	NR	NR	No	3.3	6.3	No	13.5	12	No	3.3	0	No
Median follow-up (months)	15.2	7.2	No	58	10	No	30.9	31	No	26	26	No	41	20	No

Abbreviation: NR, not reported.

* Hemoglobin drop.

adequate, and do patients have any evidence of a higher rate of recurrence? Follow-up approaching or exceeding 5 years is available from a number of centers. In none of the European or Asian series does the recurrence rate in the laparoscopically managed patients exceed that of the patients who underwent an open procedure. Spirtos *et al.*⁴¹ reported the longest experience in the United States, with recurrences being equivalent to a baseline index rate of 5% for stage I patients. Data regarding specimen size, margins, and parametria appear to be equivalent.^{36,41,43}

Finally, the literature describes 2 specifically different procedures under the heading of laparoscopic assisted radical hysterectomy: a total laparoscopic procedure that incorporates laparoscopic excision of the specimen and closure of the vagina and a procedure that involves conversion to a vaginal approach to complete the mucosal incision and closure of the vagina. Most reports from the United States favor a total laparoscopic approach, whereas the Canadian, European, and Asian series tend to incorporate some vaginal surgery.^{25,27,28,32,34–36,42,45} The value of this distinction remains to be determined but may play an important role in delineating the surgical resection margin and reducing the operative time, bladder trauma, and postoperative febrile morbidity.

ROBOTIC LAPAROSCOPIC RADICAL HYSTERECTOMY

In the last few years, telerobotic surgery has been introduced into gynecologic surgery practice. The advantages offered by this new technology include a 3-dimensional magnified field, tremor filtration, and 5 or 6 degrees of instrument mobility inside the body, thus significantly reducing the ergonomic problems associated with the conventional laparoscopic approach. There is convincing evidence that the

The advantages offered by this new technology include a 3-dimensional magnified field, tremor filtration, and 5 or 6 degrees of instrument mobility inside the body, thus significantly reducing the ergonomic problems associated with the conventional laparoscopic approach.

intuitive nature of the robotic system provides an additional advantage in terms of the learning curve.⁴⁶

Since the first robotic assisted radical hysterectomy was described in 2006,⁴⁷ a number of reports have been published, outlining advantages associated with the robotic approach.^{48–52} The use of the robot has consistently led to lower blood loss and shorter hospital stays in comparison with laparotomy. With respect to the operative time and lymph node yield, the experience has varied. Boggess *et al.*⁴⁹ was able to demonstrate a significant reduction in the operative time and increase in the number of harvested lymph nodes by using robotics in comparison with an open approach. In the series reported by Magrina *et al.*⁵² and Ko *et al.*,⁵⁰ robotic radical hysterectomy took longer with a similar nodal yield. Of major significance is the fact that all the reports were able to demonstrate a reduction in intraoperative complication rates in robotic cohorts in comparison with open cohorts (Table 2).

In the United States, only a handful of gynecologic oncology groups have had experience with using both laparoscopy and robotics for radical hysterectomy at the same institution. Our initial experience at Mount Sinai with the first 13 robotic assisted radical hysterectomies with pelvic lymphadenectomy in comparison with laparoscopic radical hysterectomy was published in 2008.⁵¹ To date, we have performed a total of 16 robotic assisted radical hysterectomies. We did not find a significant difference between the traditional laparoscopic approach and the robotic approach with respect to the operative time, operative blood loss, length of hospital stay, or complication profile. Similar findings were observed in only 2 other comparisons of robotic radical hysterectomy and laparoscopic radical hysterectomy reported by Sert *et al.*⁴⁷ (7 cases) and Magrina *et al.*⁵² (12 cases).

The advantages associated with minimally invasive approaches for patients requiring radical hysterectomy are beyond dispute. Total laparoscopic radical hysterectomy, however, remains one of the most challenging laparoscopic procedures in gynecologic oncology, requiring significant technical expertise and experience. There is no consensus on the number of cases that are required to obtain proficiency. As the number of early cervical cancer cases is decreasing, fast acquisition of advanced endoscopic skills is paramount. Therefore, the robotic interface, which allows for significant shortening of the learning curve, may make a minimally invasive approach possible even in centers with very few cases of early cervical cancer.

The robotic systems are not without drawbacks. Most commonly mentioned are the absence of tensile

Table 2. Reports of Robotic Radical Hysterectomy Versus Abdominal Radical Hysterectomy and Total Laparoscopic Hysterectomy

	Magrina <i>et al.</i> ⁵²				Ko <i>et al.</i> ⁵⁰				Boggess <i>et al.</i> ⁴⁹				Sert and Abeler ⁴⁸				Mount Sinai Series			
	Open	Laparo- scopic	Robotic	Significant Difference	Open	Robotic	Significant Difference	Open	Robotic	Open	Robotic	Significant Difference	Laparo- scopic	Robotic	Significant Difference	Laparo- scopic	Robotic	Significant Difference		
n	16	14	12	No	32	16	No	49	51	7	7	No	7	7	No	30	16	No		
Operating room time (minutes)	192.3	254.8	225.8	Yes	219	290	Yes	247.8	210.9	241	241	Yes	300	241	No	318	306	No		
Estimated blood loss (cc)	570.3	254.3	174.6	Yes	665.6	81.9	Yes	416.8	96.5	71	71	Yes	160	71	Yes	200	160	No		
Length of hospital stay	4.1	2.1	1.6	Yes	4.9	1.5	Yes	3.2	1	4	4	Yes	8	4	Yes	3.8	26	No		
Lymph node	30.6	27.9	26.2	No	17.1	15.6	No	23.3	33.8	13	13	Yes	15	13	No	31	23.3	No		
Complications	20.6 (6)	10.6 (3)	0	No	25.6 (8)	19.6 (3)	No	16.3 (6)	7.8 (6)	14 (1)	14 (1)	No	14 (1)	14 (1)	No	6.5 (2)	12 (2)	No		

feedback, the complexity, and the cost. Robotic technology is developing rapidly, however, and new instruments, smaller arms, a fourth arm, and tactile feedback should soon become available. It is true that operative expenses associated with robotic surgery currently are high; the widespread use of this technology, however, combined with shortened hospital stays hopefully will allow a substantial decrease in costs.⁵³

FERTILITY-SPARING SURGERY

Radical trachelectomy is a fertility preservation alternative to radical hysterectomy in patients with small-volume cervical cancer lesions. The cervix with adjacent parametrial tissue is resected, with the uterus left in place to allow for childbearing. Since its introduction by Dargent, Ref 14 more than 550 patients with cervical cancer have been treated with laparoscopic assisted radical vaginal trachelectomy and pelvic lymphadenectomy.^{54–56} The overall oncological safety of the operation is comparable to that of abdominal radical hysterectomy performed for lesions of similar size with a combined recurrence rate of less than 4.2% and a mortality rate of 2.8%.⁵⁷

Besides sparing fertility, laparoscopic assisted radical vaginal trachelectomy and pelvic lymphadenectomy offer multiple advantages over radical hysterectomy, including shorter hospital stays, lower blood loss, and a quicker return of bladder function.

Besides sparing fertility, laparoscopic assisted radical vaginal trachelectomy and pelvic lymphadenectomy offer multiple advantages over radical hysterectomy, including shorter hospital stays, lower blood loss, and a quicker return of bladder function.

Regardless of its potential advantages, the procedure requires advanced vaginal operative skills that are not readily learned in most gynecologic oncology training programs. Abdominal radical trachelectomy has been described, and this involves a surgical approach more familiar to most gynecologic oncologists. Einstein *et al.*⁵⁸ reported that radical abdominal trachelectomy provides similar surgical and pathological outcomes with a wider parametrial resection in comparison with the radical vaginal approach. We recently reported

a new robotically assisted approach to radical trachelectomy in the treatment of young women with cervical cancer who desire fertility preservation.⁵⁹ This technique does not require advanced vaginal surgery skills, and it combines the advantages of open and laparoscopic procedures with a familiar anatomic approach, increased magnification, and the superior dexterity of robotic surgery.

Obstetrical outcomes of radical trachelectomy are encouraging. In the most recent review,⁶⁰ close to 50% of women treated with radical trachelectomy attempted pregnancy, with over half of them

Obstetrical outcomes of radical trachelectomy are encouraging. In the most recent review, close to 50% of women treated with radical trachelectomy attempted pregnancy, with over half of them succeeding.

succeeding. The role of less radical procedures such as conization is currently under investigation.⁶¹

CONCLUSION

Abdominal radical hysterectomy at the moment remains the gold standard for the treatment of gynecologic cancers requiring a radical approach. Our newly improved understanding of the pelvic neuroanatomy has led to the development of techniques that allow for a significant reduction in the morbidity traditionally associated with radical hysterectomy. Further advances in laparoscopic and robotic technologies have allowed the safe performance of endoscopic radical hysterectomy by a greater number of surgeons along with all of the advantages associated with a minimally invasive approach. Accumulating data on radical trachelectomy give us hope for a safe fertility-sparing alternative to radical hysterectomy for women with early cervical cancer who strongly desire to bear children in the future. As we move forward, however, with our surgical and technological advances, it is imperative that oncological outcomes not be compromised and that these newer approaches be subjected to rigorous evidenced-based scrutiny before they become accepted as a new standard of care.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. *CA Cancer J Clin* 2009; 59: 225–249.
2. Schauta F. Die operation des gebaermutterkrebs mittel des schuchardtschen paravaginalschnittes. *Monatsschr Geburtshilfe Gynaekol* 1902; 15: 133–152.
3. Wertheim E. The extended abdominal operation for carcinoma uteri. *Am J Obstet Gynecol* 1912; 6: 169–232.
4. Meigs J. The Wertheim operation for carcinoma of the cervix. *Am J Obstet Gynecol* 1945; 40: 542–543.
5. Mitra S. Extraperitoneal lymphadenectomy and radical vaginal hysterectomy for cancer of the cervix (Mitra technique). *Am J Obstet Gynecol* 1959; 78: 191–196.
6. Marchiole P, Benchaib M, Buenerd A, et al. Oncological safety of laparoscopic-assisted vaginal radical trachelectomy (LARVT or Dargent's operation): a comparative study with laparoscopic-assisted vaginal radical hysterectomy (LARVH). *Gynecol Oncol* 2007; 106: 132–141.
7. Nezhat CR, Burrell MO, Nezhat FR, et al. Laparoscopic radical hysterectomy with paraaortic and pelvic node dissection. *Am J Obstet Gynecol* 1992; 166: 864–865.
8. Nezhat C. Laparoscopic radical hysterectomy with paraaortic and pelvic lymph node dissection. *Am J Obstet Gynecol* 1993; 168: 1644.
9. Canis M, Mage G, Wattiez A, et al. Does endoscopic surgery have a role in radical surgery of cancer of the cervix uteri? *J Gynecol Obstet Biol Reprod (Paris)* 1990; 19: 921.
10. Zakashansky K, Chuang L, Gretz H, et al. A case-controlled study of total laparoscopic radical hysterectomy with pelvic lymphadenectomy versus radical abdominal hysterectomy in a fellowship training program. *Int J Gynecol Cancer* 2007; 17: 1075–1082.
11. Berguer R. Surgery and ergonomics. *Arch Surg* 1999; 134: 1011–1016.
12. Dakin GF, Gagner M. Comparison of laparoscopic skills performance between standard instruments and two surgical robotic systems. *Surg Endosc* 2003; 17: 574–579.
13. US Census Bureau. Fertility of American Women: June 2004. Available at: <http://www.census.gov/prod/2005pubs/p20-555.pdf>.
14. Dargent D, Martin X, Sacchetoni A, Mathevet P. Laparoscopic vaginal radical trachelectomy: a treatment to preserve the fertility of cervical carcinoma patients. *Cancer* 2000; 88: 1877–1882.
15. Zullo MA, Mancini N, Angioli R, et al. Vesical dysfunctions after radical hysterectomy for cervical cancer: a critical review. *Crit Rev Oncol Hematol* 2003; 48: 287–293.
16. Axelsen SM, Petersen LK. Urogynaecological dysfunction after radical hysterectomy. *Eur J Surg Oncol* 2006; 32: 445–449.
17. Yabuki Y, Asamoto A, Hoshihara T, et al. Radical hysterectomy: an anatomic evaluation of parametrial dissection. *Gynecol Oncol* 2000; 77: 155–163.
18. Barton DP, Butler-Manuel SA, Buttery LD, et al. A nerve-sparing radical hysterectomy: guidelines and feasibility in Western patients. *Int J Gynecol Cancer* 2002; 12: 319–321; author reply 321.
19. Jackson KS, Naik R. Pelvic floor dysfunction and radical hysterectomy. *Int J Gynecol Cancer* 2006; 16: 354–363.
20. Charoenkwan K, Srisomboon J, Suprasert P, et al. Nerve-sparing class III radical hysterectomy: a modified technique to spare the pelvic autonomic nerves without compromising radicality. *Int J Gynecol Cancer* 2006; 16: 1705–1712.
21. Trimbois JB, Maas CP, Deruiter MC, et al. A nerve-sparing radical hysterectomy: guidelines and feasibility in Western patients. *Int J Gynecol Cancer* 2001; 11: 180–186.
22. Yabuki Y, Asamoto A, Hoshihara T, et al. A new proposal for radical hysterectomy. *Gynecol Oncol* 1996; 62: 370–378.
23. Covens A, Rosen B, Murphy J, et al. How important is removal of the parametrium at surgery for carcinoma of the cervix? *Gynecol Oncol* 2002; 84: 145–149.
24. Dottino PS, Segna RA, Zakashansky K, et al. A case controlled study of radical vaginal hysterectomy with laparoscopic lymphadenectomy versus radical abdominal hysterectomy for the treatment of early stage cervical cancer. *Gynecol Oncol* 2006; 101: 18.
25. Abu-Rustum NR, Gemignani ML, Moore K, et al. Total laparoscopic radical hysterectomy with pelvic lymphadenectomy using the argon-beam coagulator: pilot data and comparison to laparotomy. *Gynecol Oncol* 2003; 91: 402–409.
26. Coruasic A, Barisic D, Plavec A, et al. Laparoscopic surgery in cervical cancer: a Croatian experience. *Coll Antropol* 2007; 31(suppl 2): 147–154.
27. Frumovitz M, dos Reis R, Sun CC, et al. Comparison of total laparoscopic and abdominal radical hysterectomy for patients with early-stage cervical cancer. *Obstet Gynecol* 2007; 110: 96–102.
28. Ghezzi F, Cromi A, Ciravolo G, et al. Surgicopathologic outcome of laparoscopic versus open radical hysterectomy. *Gynecol Oncol* 2007; 106: 502–506.
29. Gil-Moreno A, Diaz-Feijoo B, Roca I, et al. Total laparoscopic radical hysterectomy with intraoperative sentinel node identification in patients with early invasive cervical cancer. *Gynecol Oncol* 2005; 96: 187–193.
30. Jackson KS, Das N, Naik R, et al. Laparoscopically assisted radical vaginal hysterectomy vs. radical abdominal hysterectomy for cervical cancer: a match controlled study. *Gynecol Oncol* 2004; 95: 655–661.
31. Lee CL, Huang KG. Total laparoscopic radical hysterectomy using Lee-Huang portal and McCartney transvaginal tube. *J Am Assoc Gynecol Laparosc* 2002; 9: 536–540.
32. Lee CL, Huang KG, Jain S, et al. Comparison of laparoscopic and conventional surgery in the treatment of early cervical cancer. *J Am Assoc Gynecol Laparosc* 2002; 9: 481–487.
33. Lee CL, Huang KG, Wang CJ, et al. Laparoscopic radical hysterectomy using pulsed bipolar system: comparison with conventional bipolar electrosurgery. *Gynecol Oncol* 2007; 105: 620–624.
34. Li G, Yan X, Shang H, et al. A comparison of laparoscopic radical hysterectomy and pelvic lymphadenectomy and laparotomy in the treatment of Ib-IIa cervical cancer. *Gynecol Oncol* 2007; 105: 176–180.
35. Malur S, Possover M, Schneider A. Laparoscopically assisted radical vaginal versus radical abdominal hysterectomy type II in patients with cervical cancer. *Surg Endosc* 2001; 15: 289–292.
36. Malzoni M, Malzoni C, Perone C, et al. Total laparoscopic radical hysterectomy (type III) and pelvic

- lymphadenectomy. *Eur J Gynaecol Oncol* 2004; 25: 525–527.
37. Nezhat F, Mahdavi A, Nagarsheth NP. Total laparoscopic radical hysterectomy and pelvic lymphadenectomy using harmonic shears. *J Minim Invasive Gynecol* 2006; 13: 20–25.
38. Obermair A, Ginbey P, McCartney AJ. Feasibility and safety of total laparoscopic radical hysterectomy. *J Am Assoc Gynecol Laparosc* 2003; 10: 345–349.
39. Patsner B. Radical abdominal versus laparoscopic hysterectomy for stage IB cervical cancer: what's the point? *Eur J Gynaecol Oncol* 2000; 21: 466–468.
40. Pomel C, Atallah D, Le Bouedec G, et al. Laparoscopic radical hysterectomy for invasive cervical cancer: 8-year experience of a pilot study. *Gynecol Oncol* 2003; 91: 534–539.
41. Spirtos NM, Eisenkop SM, Schlaerth JB, Ballon SC. Laparoscopic radical hysterectomy (type III) with aortic and pelvic lymphadenectomy in patients with stage I cervical cancer: surgical morbidity and intermediate follow-up. *Am J Obstet Gynecol* 2002; 187: 340–348.
42. Steed H, Rosen B, Murphy J, et al. A comparison of laparoscopic-assisted radical vaginal hysterectomy and radical abdominal hysterectomy in the treatment of cervical cancer. *Gynecol Oncol* 2004; 93: 588–593.
43. Xu H, Chen Y, Li Y, et al. Complications of laparoscopic radical hysterectomy and lymphadenectomy for invasive cervical cancer: experience based on 317 procedures. *Surg Endosc* 2007; 21: 960–964.
44. Morgan DJ, Hunter DC, McCracken G, et al. Is laparoscopically assisted radical vaginal hysterectomy for cervical carcinoma safe? A case control study with follow up. *BJOG* 2007; 114: 537–542.
45. Gil-Moreno A, Puig O, Perez-Benavente MA, et al. Total laparoscopic radical hysterectomy (type II-III) with pelvic lymphadenectomy in early invasive cervical cancer. *J Minim Invasive Gynecol* 2005; 12: 113–120.
46. Ahlering TE, Skarecky D, Lee D, Clayman RV. Successful transfer of open surgical skills to a laparoscopic environment using a robotic interface: initial experience with laparoscopic radical prostatectomy. *J Urol* 2003; 170: 1738–1741.
47. Sert BM, Abeler VM. Robotic-assisted laparoscopic radical hysterectomy (Piver type III) with pelvic node dissection—case report. *Eur J Gynaecol Oncol* 2006; 27: 531–533.
48. Sert B, Abeler V. Robotic radical hysterectomy in early-stage cervical carcinoma patients, comparing results with total laparoscopic radical hysterectomy cases. The future is now? *Int J Med Robot* 2007; 3: 224–228.
49. Boggess JF, Gehrig PA, Cantrell L, et al. A case-control study of robot-assisted type III radical hysterectomy with pelvic lymph node dissection compared with open radical hysterectomy. *Am J Obstet Gynecol* 2008; 199: 357e1–357e7.
50. Ko EM, Muto MG, Berkowitz RS, Feltmate CM. Robotic versus open radical hysterectomy: a comparative study at a single institution. *Gynecol Oncol* 2008; 111: 425–430.
51. Nezhat FR, Datta MS, Liu C, et al. Robotic radical hysterectomy versus total laparoscopic radical hysterectomy with pelvic lymphadenectomy for treatment of early cervical cancer. *JSLS* 2008; 12: 227–237.
52. Magrina JF, Kho RM, Weaver AL, et al. Robotic radical hysterectomy: comparison with laparoscopy and laparotomy. *Gynecol Oncol* 2008; 109: 86–91.
53. Burgess SV, Atug F, Castle EP, et al. Cost analysis of radical retropubic, perineal, and robotic prostatectomy. *J Endourol* 2006; 20: 827–830.
54. Abu-Rustum NR, Neubauer N, Sonoda Y, et al. Surgical and pathologic outcomes of fertility-sparing radical abdominal trachelectomy for FIGO stage IB1 cervical cancer. *Gynecol Oncol* 2008; 111: 261–264.
55. Dargent D, Mathevet P. Schauta's vaginal hysterectomy combined with laparoscopic lymphadenectomy. *Baillieres Clin Obstet Gynaecol* 1995; 9: 691–705.
56. Ramirez PT, Schmeler KM, Soliman PT, Frumovitz M. Fertility preservation in patients with early cervical cancer: radical trachelectomy. *Gynecol Oncol* 2008; 110(suppl 2): S25–S28.
57. Hertel H, Kohler C, Grund D, et al. Radical vaginal trachelectomy (RVT) combined with laparoscopic pelvic lymphadenectomy: prospective multicenter study of 100 patients with early cervical cancer. *Gynecol Oncol* 2006; 103: 506–511.
58. Einstein MH, Park KJ, Sonoda Y, et al. Radical vaginal versus abdominal trachelectomy for stage IB1 cervical cancer: a comparison of surgical and pathologic outcomes. *Gynecol Oncol* 2009; 112: 73–77.
59. Chuang LT, Lerner DL, Liu CS, Nezhat FR. Fertility-sparing robotic-assisted radical trachelectomy and bilateral pelvic lymphadenectomy in early-stage cervical cancer. *J Minim Invasive Gynecol* 2008; 15: 767–770.
60. Dornhofer N, Hockel M. New developments in the surgical therapy of cervical carcinoma. *Ann N Y Acad Sci* 2008; 1138: 233–252.
61. Rob L, Charvat M, Robova H, et al. Less radical fertility-sparing surgery than radical trachelectomy in early cervical cancer. *Int J Gynecol Cancer* 2007; 17: 304–310.

Chemotherapy for Gynecologic Cancers

Jamal Rahaman, MD, Norma Steiner, MD, Monica Prasad Hayes, MD,
Linus Chuang, MD, David Fishman, MD, and Herbert Gretz III, MD

Mount Sinai School of Medicine, New York, NY

OUTLINE

OVARIAN CANCER

Early-Stage Disease

Advanced-Stage Disease

Targeted Therapy

RECURRENT OVARIAN CANCER

Platinum-Sensitive

Recurrent Ovarian Cancer

Platinum-Resistant

Recurrent Ovarian Cancer

Neoadjuvant Chemotherapy for

Advanced Ovarian Cancer

ENDOMETRIAL CANCER AND CYTOTOXIC

CHEMOTHERAPY

Adjuvant Chemotherapy for

High-Risk Early-Stage Disease

CHEMOTHERAPY FOR

UTERINE SARCOMA

Malignant Mixed Müllerian Tumor

Leiomyosarcoma

GESTATIONAL TROPHOBLASTIC NEOPLASIA

Nonmetastatic, Low-Risk

Gestational Trophoblastic Neoplasia

Metastatic and High-Risk

Gestational Trophoblastic Neoplasia

Salvage Therapy for High-Risk

Gestational Trophoblastic Neoplasia

CERVICAL CANCER

chemotherapeutic options for early, advanced and recurrent disease are in the adjuvant setting as well as in the neoadjuvant setting are explored. For uterine cancer, adjuvant chemotherapy is employed for high risk epithelial subtypes with early disease, such as uterine papillary serous carcinomas, uterine carcinosarcomas and leiomyosarcomas, advanced stage cases, as well as recurrent disease. The review then proceeds to further discuss the appropriate treatment based on the International Federation of Gynecology and Obstetrics prognostic scoring system for gestational trophoblastic neoplasia. Finally, chemotherapy is utilized in cervical cancer as neoadjuvant therapy prior to surgery or radiation, as a sensitizer concomitantly with radiation therapy or for the treatment of advanced and recurrent disease. *Mt Sinai J Med* 76:577–588, 2009. © 2009 Mount Sinai School of Medicine

Key Words: chemotherapy, gynecological malignancies, ovarian cancer, endometrial cancer, uterine sarcoma, cervical cancer, gestational trophoblastic neoplasia, chemotherapy, bevacizumab.

The role of systemic therapy for gynecologic cancers has evolved through randomized clinical trials testing novel combination chemotherapy sequences and schedules, the introduction of novel chemotherapy agents, and the integration of emerging biologic and targeted therapies.

ABSTRACT

This review highlights significant recent developments and trends in chemotherapy for major gynecologic malignancies, i.e., ovarian cancer, endometrial cancer, uterine sarcomas, gestational trophoblastic neoplasia, and cervical cancer. In ovarian cancer,

OVARIAN CANCER

Ovarian cancer has been described as one of the most chemo-sensitive malignancies. Although there are approximately 21,550 new cases a year in the US, 14,600 deaths are expected in 2009,¹ the majority of patients who initially respond will ultimately recur and die of their disease. Approximately 70–80% of newly diagnosed patients with advanced disease exhibit a complete clinical response to primary platinum plus taxane chemotherapy.^{2,3} Carboplatin, a second generation platinum compound which is less nephrotoxic, neurotoxic and emetogenic but as efficacious as cisplatin,⁴ is now the most commonly used platinum compound. Although there is universal

Address Correspondence to:

Jamal Rahaman

Mount Sinai School of Medicine

New York, NY

Email: jamal.rahaman@mssm.edu

consensus that platinum and paclitaxel are the standard chemotherapy agents for initial treatment for ovarian cancer, questions still remain regarding the best route, dose density, dose intensity, and sequence of this combination for primary therapy.

Although there is universal consensus that platinum and paclitaxel are the standard chemotherapy agents for initial treatment for ovarian cancer, questions still remain regarding the best route, dose density, dose intensity, and sequence of this combination for primary therapy.

Additionally, the best chemotherapeutic treatment for patients with recurrence is still debated.

Early-Stage Disease

It is widely accepted that patients with a low-risk, stage IA, grade 1 tumor treated with surgery alone have an excellent prognosis, with a 95% overall 5-year survival rate, so adjuvant chemotherapy is unnecessary.^{5,6} Still, 20% to 30% of women with early-stage ovarian cancer eventually die from their disease. Therefore, it seems logical that chemotherapy should be reserved for patients at high risk for recurrence. However, defining *high risk* has been somewhat challenging.

The overall survival benefits associated with the adjuvant administration of chemotherapy in women with high-risk early-stage ovarian cancer have been documented in evidence-based trials with long-term follow-up (5–10+ years).⁷

The overall survival benefits associated with the adjuvant administration of chemotherapy in women with high-risk early-stage ovarian cancer have been documented in evidence-based trials with long-term follow-up (5–10+ years).

A combined analysis of 2 parallel randomized clinical trials studying early ovarian cancer, International Collaborative Ovarian Neoplasm 1 (ICON1)

and Adjuvant Chemotherapy in Ovarian Neoplasm (ACTION), which compared platinum-containing adjuvant chemotherapy to observation following surgery, was performed. Among the 924 patients enrolled, there was an 8% difference in survival between the cohort that received adjuvant chemotherapy and the patients that were observed.⁸ On the basis of the reduced recurrence rate and the improvement in survival, it is now recommended that patients with stage IA, IB (grade 2 or higher) and IC (all grades) should be treated with adjuvant chemotherapy.

One of the most controversial issues surrounding treatment for early-stage disease revolves around the appropriate duration of treatment.

One of the most controversial issues surrounding treatment for early-stage disease revolves around the appropriate duration of treatment.

Traditionally, chemotherapy has been given over 6 to 8 cycles in advanced ovarian cancer. Gynecologic Oncology Group (GOG) 157 was a trial of adjuvant chemotherapy randomizing patients with stage IA (grade 3), stage IB (grade 3), clear cell, stage IC, or completely resected stage II epithelial ovarian cancer to 3 cycles versus 6 cycles of paclitaxel and carboplatin. The authors concluded that after complete surgical staging, 3 cycles are reasonable for women with high-risk, early-stage epithelial ovarian cancer (EOC) and that 3 additional cycles are likely to produce only a small reduction in the risk of recurrence but with increased toxicity. However, there were several issues with this study. The regimen used carboplatin at an area-under-the-curve (AUC) of 7.5 whereas conventional therapy utilizes a lower standard dose of an AUC of 5–6. The risk of recurrence was 25.4% in the 3 cycles arm vs. 20.1% in the 6 cycles arm- and it was questioned if there were adequate patient numbers allocated in the statistical design to exclude a real difference. In addition 127 out of 427 eligible patients had incomplete staging.

Advanced-Stage Disease

Although platinum- and taxane-based chemotherapy has remained the cornerstone of treatment for over a decade, there have been many exciting advances and chemotherapeutic options in recent years. A significant modification of this standard of care has been the reintroduction of intraperitoneal chemotherapy.

Although platinum- and taxane-based chemotherapy has remained the cornerstone of treatment for over a decade, there have been many exciting advances and chemotherapeutic options in recent years. A significant modification of this standard of care has been the reintroduction of intraperitoneal chemotherapy.

There have been 3 randomized controlled trials that have shown a survival benefit in ovarian cancer patients who were administered intraperitoneal therapy instead of conventional intravenous chemotherapy.^{10–12} GOG 172 reported a 16-month survival benefit in those patients receiving intraperitoneal chemotherapy versus patients who received intravenous therapy with paclitaxel and cisplatin. By the direct delivery of cisplatin and paclitaxel as frontline therapy into the intraperitoneal cavity in patients who have undergone optimal surgical cytoreduction (defined as no residual tumor >1 cm in maximal diameter after the initial surgical cytoreduction), there is a statistically significant improvement in both progression-free and overall survival.¹⁰ The National Cancer Institute endorsed the utilization of intraperitoneal therapy in a National Cancer Institute consensus statement published in 2006 shortly after the findings of this study were published. This strategy, however, is associated with a higher risk of treatment-related side effects because of the use of cisplatin (rather than carboplatin) and the requirement for intraperitoneal catheter placement for drug delivery.¹³ Because of toxicity, only 42% of patients in the GOG 178 trial were able to complete all 6 cycles, and the trial did not use the prevailing gold standard of paclitaxel and carboplatin (GOG 158) for comparison. Because of these issues, there remains resistance to the universal application of this protocol in frontline chemotherapy; some clinicians have adopted a modification of the intraperitoneal protocol, and others continue with carboplatin and paclitaxel intravenous chemotherapy, whereas many patients enroll in clinical trials with biologic therapy.

In a preliminary study presented at the 2008 meeting of the American Society of Clinical Oncology, results of a prospective randomized trial comparing conventional dosing to a dose-dense schedule of higher doses administered over a shorter time

were reported. Patients were randomized to receive carboplatin with an area under the plasma concentration–time curve of 6 with either paclitaxel at 180 mg/m² on day 1 or paclitaxel at 80 mg/m² on days 1, 8, and 15. The treatments were repeated every 3 weeks for 6 cycles. The study reported that patients were 30% less likely to have progression of their disease if paclitaxel was administered every week versus every 21 days. These data are very intriguing; if they are confirmed, this strategy may provide responses and survival outcomes superior to those of intraperitoneal chemotherapy with lower toxicity.¹⁴

Targeted Therapy

The development of targeted molecular therapy has been in the forefront of clinical trials in the last 10 years. Although initial studies were somewhat disappointing, results of phase II trials with the anti-angiogenesis agent bevacizumab have been quite encouraging in ovarian cancer.

The development of targeted molecular therapy has been in the forefront of clinical trials in the last 10 years. Although initial studies were somewhat disappointing, results of phase II trials with the anti-angiogenesis agent bevacizumab have been quite encouraging in ovarian cancer.

Bevacizumab is thought to block angiogenesis, thereby shrinking and preventing the growth of a new tumor. An objective response rate of 20% has been seen with single-agent bevacizumab administered to patients with recurrent and resistant ovarian cancer.^{15,16} Garcia *et al.*¹⁷ demonstrated a 24% partial response and a 44% stable disease rate in a heavily pretreated group of patients with recurrent ovarian cancer who received bevacizumab and metronomic oral cyclophosphamide.

To establish the role of bevacizumab in frontline therapy for advanced disease, 2 phase III randomized prospective trials are currently in progress: GOG 218 and ICON7. Both trials include 6 cycles of standard platinum-taxane chemotherapy. The GOG 218 trial is seeking to determine not only whether bevacizumab adds any additional benefit in combination with conventional chemotherapy but also whether extended

consolidation therapy with bevacizumab is of any benefit in preventing recurrent disease. The GOG has also initiated the GOG 213 trial, which is testing the additional benefit of bevacizumab to second-line therapy with paclitaxel and carboplatin for platinum-sensitive recurrent ovarian cancer and the utility of maintenance bevacizumab.

RECURRENT OVARIAN CANCER

Despite the initial response to chemotherapy, the overwhelming majority of patients with advanced ovarian cancer will experience a recurrence. The most important factor in the further management of treatment is determining a patient's platinum sensitivity. The most widely considered definition of *platinum response* is the response to first-line platinum treatment and the length of disease-free (or platinum-free) intervals.^{18–20} Patients with relapses >6 months after the completion of first-line platinum-based chemotherapy are considered platinum-sensitive. *Platinum resistance* can be defined as any progression on treatment (also called the refractory group) or relapse within 6 months of treatment completion. There is no absolute interval that completely separates these patients. It is clear that response rates for retreatment with platinum increase over a continuum as the interval from previous therapy lengthens: platinum-free interval <6 months, 10%; 6 to 12 months, 27%; 13 to 24 months, 33%; and >24 months, 59%.¹⁸

Platinum-Sensitive Recurrent Ovarian Cancer

It is recognized that each patient must be individually assessed to determine optimal therapy in terms of recurrence, sensitivity to platinum, toxicity, ease of administration, and patient preference. All suitable patients should be offered the opportunity to participate in randomized trials if they are available. In the absence of contraindications, combination platinum-based chemotherapy should be considered for patients with prior sensitivity to platinum-containing chemotherapy.

In the absence of contraindications, combination platinum-based chemotherapy should be considered for patients with prior sensitivity to platinum-containing chemotherapy.

Two large randomized clinical trials have proved the superiority of a platinum-based doublet over platinum monotherapy in platinum-sensitive recurrent patients. The ICON4/Arbeitsgemeinschaft Gynäkologische Onkologie Studiengruppe Ovarialkarzinom 2.2 (AGO-OVAR 2.2) trial demonstrated an improved survival benefit for patients who received a combination of paclitaxel and carboplatin (or cisplatin) compared to those who received carboplatin monotherapy.²¹ This benefit was clearer in patients with a treatment-free interval >12 months. Moreover, the AGO-OVAR 2.5 trial, with the cooperation of the National Cancer Institute of Canada Clinical Trials Group (NCIC-CTG) and the European Organisation for Research and Treatment of Cancer Gynaecological Cancer Group (EORTC-GRG), has confirmed the advantage in the response rate and progression-free survival of the carboplatin-gemcitabine doublet in comparison with carboplatin.²² The Caelyx in Platinum Sensitive Ovarian Patients (CALYPSO) trial was presented at ASCO 2009 and was a multicenter phase III study designed to compare efficacy and safety of carboplatin-pegylated liposomal doxorubicin (PLD) (C-D) and carboplatin-paclitaxel (C-P) in relapsed platinum-sensitive Ovarian Cancer patients. This trial, the largest in relapsed ovarian cancer (976 enrolled patients), showed significant superiority of PLD-carboplatin combination in terms of PFS (11.3 vs. 9.4 months, HR 0.821, $P = 0.005$). In addition, compared to paclitaxel-carboplatin, PLD-carboplatin was well tolerated with lower rates of severe and long-lasting (neuropathy) toxicities.

If combination platinum-based chemotherapy is not appropriate, then a single platinum agent should be considered. Carboplatin has demonstrated efficacy across trials and has a manageable toxicity profile. If a single platinum agent is not being considered, then monotherapy with paclitaxel, topotecan, gemcitabine or pegylated liposomal doxorubicin are seen as reasonable treatment options. Some patients may be repeatedly sensitive to treatment and may benefit from multiple lines of chemotherapy.

Platinum-Resistant Recurrent Ovarian Cancer

The issue of combination chemotherapy versus monotherapy is a clinically relevant dilemma for medical and gynecologic oncologists involved in the treatment of recurrent patients. In those patients with platinum-resistant relapse, combination chemotherapy has been associated with higher toxicity without a clear clinical benefit in randomized clinical

trials. Therefore, enrollment in clinical trials or a less toxic sequential monotherapy approach should be the choice for resistant patients. The list of agents active in ovarian cancer includes cisplatin, carboplatin, paclitaxel, docetaxel, topotecan, doxorubicin, pegylated liposomal doxorubicin, gemcitabine, cyclophosphamide, etoposide, ifosfamide, vinorelbine, altretamine (hexamethylmelamine) and melphalan (alkeran). In addition premetrexed and trabectedin have recently been shown to have significant activity in ovarian cancer.

Neoadjuvant Chemotherapy for Advanced Ovarian Cancer

Finally, neoadjuvant chemotherapy prior to debulking surgery is the subject of several ongoing clinical trials and may turn out to be one of the most important developments since the concept of interval debulking surgery was established and proven in Europe.

Neoadjuvant chemotherapy prior to debulking surgery is the subject of several ongoing clinical trials and may turn out to be one of the most important developments since the concept of interval debulking surgery was established and proven in Europe.

No prospective randomized trials in the United States have evaluated neoadjuvant chemotherapy against conventional treatment for advanced ovarian cancer. One international prospective randomized trial conducted by the European Organization for Research and Treatment of Cancer (EORTC) was recently presented at the 2008 IGCS Meeting. The EORTC 55791 trial randomized 718 women with Stage IIIC-IV ovarian cancer to receive either neoadjuvant chemotherapy for 3 cycles followed by cytoreductive surgery and then 3 additional cycles of chemotherapy, or surgery followed by 6 cycles of chemotherapy. For the latter arm, there was an option to perform interval cytoreduction after 3 cycles of chemotherapy if the surgeon declared that it was his or her intent to do so prior to performing the initial surgery. Because there was no difference in PFS or survival between conventional therapy and neoadjuvant chemotherapy in the EORTC data; and the morbidity was improved among women receiving neoadjuvant chemotherapy, neoadjuvant

chemotherapy may become an accepted standard technique for the management of patients with advanced-stage ovarian cancer.²³ A very similar trial, Chemotherapy or Upfront Surgery (CHORUS), is now accruing patients in a UK study with a target accrual of 550 patients. Finally, the Japanese Clinical Oncology Group is inviting patients to participate in a prospective randomized trial in which 350 women will be randomized to receive either 4 cycles of carboplatin and paclitaxel followed by surgery and 4 additional cycles of treatment or initial cytoreductive surgery followed by 8 cycles of carboplatin and paclitaxel.²⁴

ENDOMETRIAL CANCER AND CYTOTOXIC CHEMOTHERAPY

Endometrial cancer is the most common of the gynecologic cancers, with 42,160 new cases expected in 2009.¹ The majority of endometrial adenocarcinomas are detected at an early stage, with the tumor confined to the uterine corpus, and thus the prognosis is favorable, with a 5-year survival rate of 90% for stage I and a 5-year survival rate of 83% for stage II.²⁵ However, there are still over 7700 deaths from endometrial cancer each year in the United States, and adjuvant chemotherapy is employed for high-risk subtypes with early disease, advanced-stage cases, and recurrent disease. The 4 most active agents for endometrial cancer are doxorubicin, cisplatin, carboplatin, and paclitaxel. The GOG has conducted a series of randomized phase III clinical trials to define the most effective combination-chemotherapy regimen for endometrial cancer.

The GOG studied single-agent doxorubicin at a dose rate of 60 mg/m² intravenously every 3 weeks and reported a 37% response rate in 43 patients.²⁶ The most frequent adverse effects were myelosuppression and cardiac toxicity. This study was important because it clearly established the value of doxorubicin as a single agent in the chemotherapy of endometrial cancer.

A combination of cisplatin and doxorubicin in advanced or recurrent endometrial cancer produced response rates of 33% to 80%, which depended on the proportion of patients previously treated with either chemotherapy or radiation.^{27,28} In the EORTC study reported by Aapro and colleagues,²⁹ a survival advantage was demonstrated with a response rate of 43% for the doxorubicin and cisplatin combination versus 17% for single-agent doxorubicin ($P < 0.001$).

The most recent completed phase III prospective randomized GOG study (Protocol 177) compared

standard doxorubicin and cisplatin (AP) against paclitaxel, doxorubicin, and cisplatin (TAP).³⁰ Two hundred sixty-six patients were randomized, and the results indicated that TAP produced a significant improvement in the response rate (57% versus 34%; $P < 0.01$), progression-free survival (median: 8.3 versus 5.3 months; $P < 0.01$), and overall survival (median: 15.3 versus 12.3 months; $P = 0.037$).^{31,32} Neurologic toxicity was worse for those receiving TAP, with 12% grade 3, and 27% grade 2 peripheral neuropathy, compared with 1% and 4%, respectively, in those receiving AP. The GOG is currently randomizing patients with advanced or recurrent disease to either TAP or a combination of paclitaxel and carboplatin, which is easy to administer and has an improved toxicity profile (GOG Protocol 209).

Adjuvant Chemotherapy for High-Risk Early-Stage Disease

High-risk endometrial cancers such as uterine papillary serous carcinoma (UPSC) and clear cell carcinomas account for 10% to 15% of uterine endometrial carcinomas and are aggressive tumors with a poor prognosis and high recurrence rates even in early stages. Thus, these patients have been treated with a combination of modalities in multiple retrospective studies, and the data suggest that all patients with these diagnoses should be treated with systemic chemotherapy with or without radiotherapy. However, there is still no randomized phase III trial data to support the use of chemotherapy in high risk early stage endometrial cancer.

In a study of 74 patients with stage I UPSC, Kelly *et al.*³³ demonstrated that all patients with residual disease (including disease confined to the endometrium) had improved progression-free survival and overall survival with adjuvant platinum-based chemotherapy. There were no vaginal recurrences in those receiving vaginal brachytherapy. Because clear cell carcinoma behaves like UPSC, despite the absence of data, they are currently treated in a similar fashion. The GOG 249 was opened in March 2009 and is a phase III trial of pelvic Radiation versus vaginal cuff brachytherapy followed by 3 cycles of chemotherapy with carboplatin (AUC = 6) and paclitaxel (175 mg/m²) in patients with high risk early-stage endometrial cancer.

CHEMOTHERAPY FOR UTERINE SARCOMA

Uterine sarcomas, namely leiomyosarcoma (LMS) and malignant mixed müllerian tumor (MMMT), although rare, are more lethal than their epithelial counterparts with high recurrence rates of at least 50% even in stage I disease. Because of the propensity for distant recurrence, the need for effective systemic adjuvant therapy is clear; however, it is very difficult to evaluate existing chemotherapy agents or assess novel compounds in these tumor types because of the long accrual times for clinical trials of these rare entities.

Malignant Mixed Müllerian Tumor

Ifosfamide is the most active single agent in the treatment of advanced or recurrent MMMT of the uterus. Sutton and colleagues³⁴ conducted a phase II study of ifosfamide and mesna in patients with advanced or recurrent MMMT and found a 32% overall response rate. However, the response duration ranged from 1.4 to 8.6 months, with a median response duration of only 3.8 months.

A phase III trial (GOG Protocol 108) compared ifosfamide with and without cisplatin in patients with advanced or recurrent MMMT.³⁵ Overall, 54% of the patients treated with the combination regimen and 36% of the patients treated with ifosfamide alone responded to therapy. This combination offered a slight prolongation of progression-free survival but no significant overall survival benefit.

The most recent phase III trial compared ifosfamide with and without paclitaxel. The crude response rate was 29% with ifosfamide alone and 45% with the combination therapy. There was a 31% decrease in the hazard of death (hazard ratio, 0.69; 95% confidence interval, 0.49–0.97; $P = 0.03$) as well as a 29% decrease in the hazard of progression (hazard ratio, 0.71; 95% confidence interval, 0.51–0.97; $P = 0.03$) with combination chemotherapy versus ifosfamide alone. Toxicities in the combination arm were expected and manageable.³⁶

Toyoshima *et al.*³⁷ recently reported on six patients with advanced or recurrent uterine carcinosarcoma treated with carboplatin (AUC = 6) plus paclitaxel (175 mg/m²). Four of the five patients with measurable disease demonstrated complete responses. Toxicity was relatively minimal and felt to be less than traditional regimens used in this tumor type. The GOG 232B is a phase II trial which completed accrual in March 2008 and was designed to assess the activity of carboplatin (AUC = 6) plus

Table 1. 2000 Revised FIGO Prognostic Scoring System for GTN⁴¹.

FIGO Score	0	1	2	3
Age	<39	>39	—	
Antecedent Pregnancy	Mole	Abortion	Term pregnancy	
Interval Months from Index Pregnancy	<4	4–6	7–12	>12
Pretreatment hCG (IU/L)	<10 ³	10 ³ –10 ⁴	>10 ⁴ –10 ⁵	>10 ⁵
Largest Tumor, including Uterus	3–4 cm	5 cm		
Sites of metastasis		Spleen/Kidney	GI tract	Liver/Brain
Number of metastasis	0	1–4	≥8	>8
Previously failed chemotherapy			Single drug	2 or more drugs

paclitaxel (175 mg/m²) in patients with recurrent or advanced uterine carcinosarcomas (MMMT). This data are not mature and have not yet been presented.

Leiomyosarcoma

Of the single agents that have been tested in patients with LMS, doxorubicin is the most active, with a response rate of up to 25%.³⁸ Liposomal doxorubicin, however, had a disappointing response rate of only 16%.³⁹ Ifosfamide exhibited moderate activity (17.2%).⁴⁰ Single-agent gemcitabine had a 20.5% response rate. The GOG conducted a phase II study of doxorubicin in combination with ifosfamide-mesna in previously untreated patients with advanced or recurrent uterine LMS.⁴¹ Although the overall response rate was 33.3%, the toxicity was severe and included grade 4 neutropenia and grade 4 cardiotoxicity.

After an initial single-institution report by Hensley *et al.*,⁴² the GOG conducted 2 phase II studies examining the combination of gemcitabine and docetaxel in metastatic LMS. When the combination was used as first-line therapy, objective responses were observed in 15 of 42 patients (overall, 35.8%; complete response, 4.8%; partial response, 31%; 90% confidence interval, 23.5%–49.6%), with an additional 11 (26.2%) having stable disease. The median progression-free survival was 4.4 months (range, 0.4–37.2+ months). Among 15 women with an objective response, the median response duration was 6 months (range, 2.1–33.4+ months). The median overall survival was 16+ months (range, 4–41.3 months). The authors concluded that fixed-dose-rate gemcitabine and docetaxel constitute a reasonable option for first-line treatment of uterine LMS. Among 48 evaluable patients in the second-line GOG trial, the overall objective response rate was 27%, with a complete response in 6.3% (3/48) and a partial response in 20.8% (10/48). An

additional 50% (24/48) had stable disease (median duration, 5.4 months).

GESTATIONAL TROPHOBLASTIC NEOPLASIA

Gestational trophoblastic neoplasia (GTN) includes invasive mole, choriocarcinoma, and placental site trophoblastic tumors. It is one of the most curable solid tumors, with an overall cure rate exceeding 90%.⁴³ Thorough evaluation and staging allow the selection of appropriate therapy while minimizing toxicity. In 2000, the International Federation of Gynecology and Obstetrics (FIGO) revised their staging system and modified the previous World Health Organization (WHO) prognostic scoring system to distinguish nonmetastatic and low-risk metastatic trophoblastic tumors from high-risk metastatic tumors on the basis of their predictable resistance to methotrexate (Table 1).⁴⁴ Each prognostic value is assigned a numerical number and is then tallied. Nonmetastatic GTN and low-risk metastatic GTN (revised FIGO score ≤6) can be treated with single-agent chemotherapy, resulting in a survival rate approaching 100%. High-risk metastatic GTN (FIGO ≥7) requires initial multiagent chemotherapy with or without adjuvant radiation and surgery to achieve survival rates of 80% to 90%.

Nonmetastatic, Low-Risk Gestational Trophoblastic Neoplasia

Patients with nonmetastatic and low-risk metastatic GTN have an excellent prognosis and therefore may be treated with single-agent chemotherapy. The most widely used chemotherapy drugs are methotrexate and actinomycin D. Historically, because the short-term toxicity of actinomycin D is considered greater than that of methotrexate, methotrexate has been the more commonly used agent. Many different protocols have been devised with different

routes of administration, doses, and frequency for each chemotherapeutic agent;^{45–48} however, few randomized controlled trials have been performed to determine which regimen is best. One of the earliest regimens required daily intramuscular methotrexate injections over the course of 5 days. The reported response rates varied from 81% to 92.3%⁴⁹; however, it was associated with higher adverse events. Alternative schedules include weekly methotrexate pulse therapy with doses ranging from 30 to 50 mg/m² or methotrexate (1 mg/kg/day) alternating with folinic acid (1 mg/kg/day), each for 4 days. Another high-dose regimen includes 100 mg/m² methotrexate intravenously infused for 30 minutes followed by 200 mg/m² methotrexate in a continuous injection over 12 hours with folinic acid rescue was described by the New England Trophoblastic Disease Center Group, with complete remission reported in 44.8% to 81.5% of patients after the first chemotherapy course.⁵⁰

Recently, 2 randomized trials compared methotrexate and actinomycin D in low-risk GTN. The Gynecologic Oncology Group investigated the efficacy of pulsed actinomycin D (1.25 mg/m²) given biweekly versus parenteral methotrexate (30 mg/m²) given weekly. The remission rate with actinomycin D (69%) was significantly greater than that with methotrexate (53%).⁵¹ Yarandi *et al.*⁵² reported similar results, with complete remission achieved in 48.14% of patients in the methotrexate group and in 90% of patients in the actinomycin D group.⁵² On the basis of these results, there will perhaps be a shift toward treatment with actinomycin D.

Metastatic and High-Risk Gestational Trophoblastic Neoplasia

Patients with a FIGO score greater than 7 are at high risk for resistance to single-agent therapy and thus should be treated with combination chemotherapy.

Patients with a FIGO score greater than 7 are at high risk for resistance to single-agent therapy and thus should be treated with combination chemotherapy.

Etoposide, methotrexate, actinomycin D, cyclophosphamide, and vincristine (EMA/CO) form the most widely used combination and induce complete remission in 76% to 97% of patients^{53–55}. Various combination chemotherapies have been used in the treatment of metastatic gestational trophoblastic neoplasia. One

retrospective study compared treatment results in 227 patients with metastatic high-risk GTN who received combination chemotherapy: 49 were treated with methotrexate plus folinic acid and dactinomycin (MTX+Act-D); 40 were treated with methotrexate, actinomycin D, and cyclophosphamide (MAC); 42 were treated with cytoxan, hydroxyurea, actinomycin D, methotrexate, oncovin, calcium folinate, and Adriamycin (CHAMOCA); and 96 were treated with EMACO. The EMACO regimen resulted in complete remission in 90% of patients. Complete remission with the other regimens (MTX+Act-D, MAC, and CHAMOCA) occurred in 63%, 67%, and 76% of patients, respectively. The EMACO regimen was recorded to have few toxic effects, early remission, and a low failure rate and was considered to be the most effective regimen.⁵⁶

Despite achieving complete serological remission with single-agent chemotherapy for 6 weeks, patient may experience relapse of the disease. The relapse is defined as having 2 rising serum human chorionic gonadotropin (HCG) titers in the absence of a normal pregnancy. These patients are also best treated with EMACO and have a 100% overall survival rate.⁵⁷

Salvage Therapy for High-Risk Gestational Trophoblastic Neoplasia

For women who fail EMA/CO, salvage chemotherapy with EMA/EP with or without surgery salvages a further 75% of cases. In the EMA/EP regimen, cyclophosphamide and vincristine are replaced by etoposide (100 mg/m²) and cisplatin (60 mg/m²) on day 8. Because of the toxicity associated with EMA/CO, paclitaxel and cisplatin, alternating with etoposide and paclitaxel, have also been used. The reported overall survival is 75%. However, in patients who already failed previous chemotherapy, the overall survival dropped to 44%.⁵⁸

In summary, GTN is a rare neoplasm and disease that is highly curable with chemotherapy. Emerging evidence suggests that single-agent actinomycin D may be used to treat low-risk GTN, whereas combination therapy with EMACO remains the treatment of choice for high-risk GTN, although no prospective trials have yet been designed to confirm this practice.

CERVICAL CANCER

Chemotherapy is used in cervical cancer as neoadjuvant therapy prior to surgery or radiation, as a

sensitizer concomitantly with radiation therapy, or for the treatment of advanced and recurrent disease.

Chemotherapy became a critical component of primary radiation therapy for cervical cancer in 1999 when 5 randomized control trials showed that the addition of cisplatin chemotherapy sensitization to traditional pelvic radiotherapy provided a 30% to 50% improvement in overall survival over radiotherapy alone.^{59–63} As a result of these trials, the National Cancer Institute published a statement recommending that all cervical cancer patients treated with radiation therapy also receive cisplatin-based chemosensitization. With the advent of biologic agents, there has been much interest in combining the current chemoradiation standard with a biologic agent in order to further improve response and survival rates in locally advanced cervical carcinoma. Two such agents are being studied in phase I and II trials by the Radiation Therapy Oncology Group (RTOG) and GOG : Bevacizumab(Avastin), a humanized monoclonal antibody directed against various isoforms of vascular endothelial growth factor receptors, and cetuximab (Erbixux), a chimerized antibody specific for binding to epidermal growth factor receptor and its heterodimers.

In patients with advanced and recurrent cervical cancer for which curative therapy with surgery and/or chemoradiation is not an option, systemic chemotherapy is the standard treatment option. The single most active agent that has been studied in this patient population is cisplatin, with response rates of 13% to 38%.^{64–67} Various other agents, including liposomal doxorubicin (Doxil), vinorelbine, ifosfamide, paclitaxel, and topotecan, have been studied in phase II trials and have shown response rates of 11% to 18%.^{68–73} The 3 agents that have been studied in combination with cisplatin are paclitaxel, ifosfamide, and topotecan.^{64–66} In these trials, only the combination of topotecan and cisplatin has been shown to have an overall survival benefit over cisplatin alone of 9.4 versus 6.5 months ($P = 0.017$), with a response rate of 27% versus 13%.⁶⁴ In response to the higher response rates of combination therapy, a phase III randomized control trial (GOG 204) comparing paclitaxel and cisplatin to combinations of gemcitabine, vinorelbine, or topotecan with cisplatin was conducted by the GOG. The results, which were presented at the 2008 meeting of the American Society of Clinical Oncology, showed no superiority of any of the combinations over paclitaxel and cisplatin.⁷⁴

In the recurrent and advanced setting, there is also interest in combining biologic agents with traditional chemotherapy in order to improve outcomes. One retrospective study of a cohort of

6 heavily pretreated patients with recurrent disease receiving bevacizumab in combination with 5-fluorouracil or capecitabine showed clinical benefit in 67% of the patients and a median time to progression of 4.3 months. The GOG recently presented its phase II trial of single-agent bevacizumab in recurrent squamous cell cervical cancer, showing an 11% response rate and progression-free survival of 24% for 6 months or longer.⁷⁵ In April 2009, the GOG started accrual in its latest phase III randomized control trial (GOG 240) for advanced and recurrent cervical cancer with 4 arms : paclitaxel and cisplatin with or without bevacizumab versus topotecan and paclitaxel with or without bevacizumab.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2009. *CA Cancer J Clin* 2009; 59: 225–249.
2. McGuire WP, Hoskins WJ, Brady MF, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *N Engl J Med* 1996; 334: 1–6.
3. Covens A, Carey M, Bryson P, et al. Systematic review of first-line chemotherapy for newly diagnosed postoperative patients with stage II, III, or IV epithelial ovarian cancer. *Gynecol Oncol* 2002; 85: 71–80.
4. Ozols RF, Bundy BN, Greer BE, et al. Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. *J Clin Oncol* 2003; 21: 3194–3200.
5. Morice P, Wicart-Poque F, Rey A, et al. Results of conservative treatment in epithelial ovarian carcinoma. *Cancer* 2001; 92: 2412–2418.
6. Young RC, Walton LA, Ellenberg SS, et al. Adjuvant therapy in stage I and stage II epithelial ovarian cancer. Results of two prospective randomized trials. *N Engl J Med* 1990; 322: 1021–1027.
7. Swart AM, Burdett S, Ledermann J, et al. Why i.p. therapy cannot yet be considered as a standard of care for the first-line treatment of ovarian cancer: a systematic review. *Ann Oncol* 2008; 19: 688–695.
8. Colombo N, Pecorelli S. What have we learned from ICON1 and ACTION? *Int J Gynecol Cancer* 2003; 13(suppl 2): 140–143.
9. Bell J, Brady MF, Young RC, et al. Randomized phase III trial of three versus six cycles of adjuvant carboplatin and paclitaxel in early stage epithelial ovarian carcinoma: a Gynecologic Oncology Group study. *Gynecol Oncol* 2006; 102: 432–439.
10. Armstrong DK, Bundy B, Wenzel L, et al. Intraperitoneal cisplatin and paclitaxel in ovarian cancer. *N Engl J Med* 2006; 354: 34–43.
11. Elit L, Oliver TK, Covens A, et al. Intraperitoneal chemotherapy in the first-line treatment of women

- with stage III epithelial ovarian cancer: a systematic review with metaanalyses. *Cancer* 2007; 109: 692–702.
12. Markman M, Bundy BN, Alberts DS, et al. Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian carcinoma: an intergroup study of the Gynecologic Oncology Group, Southwestern Oncology Group, and Eastern Cooperative Oncology Group. *J Clin Oncol* 2001; 19: 1001–1007.
 13. Markman M, Walker JL. Intraperitoneal chemotherapy of ovarian cancer: a review, with a focus on practical aspects of treatment. *J Clin Oncol* 2006; 24: 988–994.
 14. Isonishi S, Yasuda M, Takahashi F, et al. Dose dense paclitaxel once a week and conventional carboplatin every 3 weeks for advanced ovarian cancer: a phase 3 open label randomised controlled trial. *Lancet* 2009; 374: 1331–1338.
 15. Burger RA, Sill MW, Monk BJ, et al. Phase II trial of bevacizumab in persistent or recurrent epithelial ovarian cancer or primary peritoneal cancer: a Gynecologic Oncology Group Study. *J Clin Oncol* 2007; 25: 5165–5171.
 16. Cannistra SA, Matulonis UA, Penson RT, et al. Phase II study of bevacizumab in patients with platinum-resistant ovarian cancer or peritoneal serous cancer. *J Clin Oncol* 2007; 25: 5180–5186.
 17. Garcia AA, Hirte H, Fleming G, et al. Phase II clinical trial of bevacizumab and low-dose metronomic oral cyclophosphamide in recurrent ovarian cancer: a trial of the California, Chicago, and Princess Margaret Hospital phase II consortia. *J Clin Oncol* 2008; 26: 76–82.
 18. Blackledge G, Lawton F, Redman C, Kelly K. Response of patients in phase II studies of chemotherapy in ovarian cancer: implications for patient treatment and the design of phase II trials. *Br J Cancer* 1989; 59: 650–653.
 19. Markman M, Hoskins W. Responses to salvage chemotherapy in ovarian cancer: a critical need for precise definitions of the treated population. *J Clin Oncol* 1992; 10: 513–514.
 20. Markman M, Rothman R, Hakes T, et al. Second-line platinum therapy in patients with ovarian cancer previously treated with cisplatin. *J Clin Oncol* 1991; 9: 389–393.
 21. Parmar MK, Ledermann JA, Colombo N, et al. Paclitaxel plus platinum-based chemotherapy versus conventional platinum-based chemotherapy in women with relapsed ovarian cancer: the ICON4/AGO-OVAR-2.2 trial. *Lancet* 2003; 361: 2099–2106.
 22. Pfisterer J, Plante M, Vergote I, et al. Gemcitabine plus carboplatin compared with carboplatin in patients with platinum sensitive recurrent ovarian cancer: An intergroup trial of the AGO OVAR, the NCIC CTG and the EORTC GCG. *J Clin Oncol* 2006; 24: 4699–4707.
 23. Schwartz PE. What is the role of neoadjuvant chemotherapy in the management of ovarian cancer? *Oncology* 2008; 22: 1118–1125.
 24. Vergote I. EORTC-GCG/NCIC-CTG Randomized Trial comparing primary debulking surgery with neoadjuvant chemotherapy in Stage IIIC-IV Ovarian, Fallopian tube and Peritoneal Cancer (OVCA). Proceedings of the International Gynecologic Cancer Society 11th Biannual Meeting 2008. Abstract #1767.
 25. Maneschi M, Maneschi F, Geraci P, et al. Surgical pathological staging of endometrial carcinoma and results of treatment. *Eur J Gynaecol Oncol* 1992; 13: 30–35.
 26. Thigpen JT, Buchsbaum HJ, Mangan C, Blessing JA. Phase II trial of Adriamycin in the treatment of advanced or recurrent endometrial carcinoma: a Gynecologic Oncology Group study. *Cancer Treat Rep* 1979; 63: 21–27.
 27. Seltzer V, Vogl SE, Kaplan BH. Adriamycin and cis-diamminedichloroplatinum in the treatment of metastatic endometrial adenocarcinoma. *Gynecol Oncol* 1984; 19: 308–313.
 28. Pasmantier MW, Coleman M, Silver RT, et al. Treatment of advanced endometrial carcinoma with doxorubicin and cisplatin: effects on both untreated and previously treated patients. *Cancer Treat Rep* 1985; 69: 539–542.
 29. Aapro MS, van Wijk FH, Bolis G, et al. Doxorubicin versus doxorubicin and cisplatin in endometrial carcinoma: definitive results of a randomised study (55872) by the EORTC Gynaecological Cancer Group. *Ann Oncol* 2003; 14: 441–448.
 30. Scarabelli C, Campagnutta E, Giorda G, et al. Maximal cytoreductive surgery as a reasonable therapeutic alternative for recurrent endometrial carcinoma. *Gynecol Oncol* 1998; 70: 90–93.
 31. Fleming GF, Filiaci VL, Bentley RC, et al. Phase III randomized trial of doxorubicin + cisplatin versus doxorubicin + 24-h paclitaxel + filgrastim in endometrial carcinoma: a Gynecologic Oncology Group study. *Ann Oncol* 2004; 15: 1173–1178.
 32. Flemming GF, Brunetto AJ, Mundt RT. Randomized trial of doxorubicin (DOX) plus cisplatin (CIS) versus DOX plus CIS plus paclitaxel (TAX) in patients with advanced or recurrent endometrial carcinoma: a Gynecologic Oncology Group (GOG) study [abstract]. *J Clin Oncol* 2002; 21: 2029.
 33. Kelly MG, O'Malley DM, Hui P, et al. Improved survival in surgical stage I patients with uterine papillary serous carcinoma (UPSC) treated with adjuvant platinum-based chemotherapy. *Gynecol Oncol* 2005; 98: 353–359.
 34. Sutton GP, Blessing JA, Homesley HD, Malfetano JH. A phase II trial of ifosfamide and mesna in patients with advanced or recurrent mixed mesodermal tumors of the ovary previously treated with platinum-based chemotherapy: a Gynecologic Oncology Group study. *Gynecol Oncol* 1994; 53: 24–26.
 35. Ramondetta LM, Burke TW, Jhingran A, et al. A phase II trial of cisplatin, ifosfamide, and mesna in patients with advanced or recurrent uterine malignant mixed mullerian tumors with evaluation of potential molecular targets. *Gynecol Oncol* 2003; 90: 529–536.
 36. Homesley HD, Filiaci V, Markman M, et al. Phase III trial of ifosfamide with or without paclitaxel in advanced uterine carcinosarcoma: a Gynecologic Oncology Group Study. *J Clin Oncol* 2007; 25: 526–531.
 37. Toyoshima M, Akahira J, Matsunaga G, et al. Clinical experience with combination paclitaxel and carboplatin therapy for advanced or recurrent carcinosarcoma of the uterus. *Gynecol Oncol* 2004; 94: 774–778.
 38. Omura GA, Major FJ, Blessing JA, et al. A randomized study of Adriamycin with and without dimethyl triazenoimidazole carboxamide in advanced uterine sarcomas. *Cancer* 1983; 52: 626–632.

39. Sutton G, Blessing J, Hanjani P, Kramer P. Phase II evaluation of liposomal doxorubicin (Doxil) in recurrent or advanced leiomyosarcoma of the uterus: a Gynecologic Oncology Group study. *Gynecol Oncol* 2005; 96: 749–752.
40. Wilson TO, Podratz KC, Gaffey TA, et al. Evaluation of unfavorable histologic subtypes in endometrial adenocarcinoma. *Am J Obstet Gynecol* 1990; 162: 418–423; discussion 423–426.
41. Sutton G, Blessing JA, Malfetano JH. Ifosfamide and doxorubicin in the treatment of advanced leiomyosarcomas of the uterus: a Gynecologic Oncology Group study. *Gynecol Oncol* 1996; 62: 226–229.
42. Hensley ML, Blessing JA, Mannel R, Rose PG. Fixed-dose rate gemcitabine plus docetaxel as first-line therapy for metastatic uterine leiomyosarcoma: a Gynecologic Oncology Group phase II trial. *Gynecol Oncol* 2008; 109: 329–334.
43. Soper JT. Gestational trophoblastic disease. *Obstet Gynecol* 2006; 108: 176–187.
44. Kohorn EI. The new FIGO 2000 staging and risk factor scoring system for gestational trophoblastic disease: description and clinical assessment. *Int J Gynecol Cancer* 2001; 11: 73–77.
45. Kwon JS, Elit L, Mazurka J, et al. Weekly intravenous methotrexate with folinic acid for nonmetastatic gestational trophoblastic neoplasia. *Gynecol Oncol* 2001; 82: 367–370.
46. Kohorn EI. Is lack of response to single-agent chemotherapy in gestational trophoblastic disease associated with dose scheduling or chemotherapy resistance? *Gynecol Oncol* 2002; 85: 36–39.
47. Foulmann K, Guastalla JP, Caminet N, et al. What is the best protocol of single-agent methotrexate chemotherapy in nonmetastatic or low-risk metastatic gestational trophoblastic tumors? A review of the evidence. *Gynecol Oncol* 2006; 102: 103–110.
48. McNeish IA, Strickland S, Holden L, et al. Low-risk persistent gestational trophoblastic disease: outcome after initial treatment with low-dose methotrexate and folinic acid from 1992 to 2000. *J Clin Oncol* 2002; 20: 1838–1844.
49. Hammond CB, Hertz R, Ross GT, et al. Primary chemotherapy for nonmetastatic gestational trophoblastic neoplasms. *Am J Obstet Gynecol* 1967; 98: 71–78.
50. Schorge JO, Lea JS, Farrar DF, et al. Management of low-risk gestational trophoblastic neoplasia in indigent women. *J Reprod Med* 2003; 48: 780–784.
51. Osborne R, Filiaci V, Schink J. A randomized phase III trial comparing weekly parenteral methotrexate and pulsed dactinomycin as primary management of low-risk gestational trophoblastic neoplasia: a Gynecologic Oncology Group Study. *Gynecol Oncol* 2008; 108: S2.
52. Yarandi F, Eftekhari Z, Shojaei H, et al. Pulse methotrexate versus pulse actinomycin D in the treatment of low-risk gestational trophoblastic neoplasia. *Int J Gynaecol Obstet* 2008; 103: 33–37.
53. Bolis G, Bonazzi C, Landoni F, et al. EMA/CO regimen in high-risk gestational trophoblastic tumor (GTT). *Gynecol Oncol* 1988; 31: 439–444.
54. Berkowitz RS, Goldstein DP. Current management of gestational trophoblastic diseases. *Gynecol Oncol* 2009; 112: 654–662.
55. Bower M, Newlands ES, Holden L, et al. EMA/CO for high-risk gestational trophoblastic tumors: results from a cohort of 272 patients. *J Clin Oncol* 1997; 15: 2636–2643.
56. Kim SJ, Bae SN, Kim JH, et al. Effects of multiagent chemotherapy and independent risk factors in the treatment of high-risk GTT—25 years experiences of KRI-TRD. *Int J Gynaecol Obstet* 1998; 60(suppl 1): S85–S96.
57. Powles T, Savage PM, Stebbing J, et al. A comparison of patients with relapsed and chemo-refractory gestational trophoblastic neoplasia. *Br J Cancer* 2007; 96: 732–737.
58. Wang J, Short D, Sebire NJ, et al. Salvage chemotherapy of relapsed or high-risk gestational trophoblastic neoplasia (GTN) with paclitaxel/cisplatin alternating with paclitaxel/etoposide (TP/TE). *Ann Oncol* 2008; 19: 1578–1583.
59. Peters WA III, Liu PY, Barrett RJ II, et al. Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. *J Clin Oncol* 2000; 18: 1606–1613.
60. Morris M, Eifel PJ, Lu J, et al. Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. *N Engl J Med* 1999; 340: 1137–1143.
61. Rose PG, Bundy BN, Watkins EB, et al. Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. *N Engl J Med* 1999; 340: 1144–1153.
62. Keys HM, Bundy BN, Stehman FB, et al. Cisplatin, radiation, and adjuvant hysterectomy compared with radiation and adjuvant hysterectomy for bulky stage IB cervical carcinoma. *N Engl J Med* 1999; 340: 1154–1161.
63. Whitney CW, Sause W, Bundy BN, et al. Randomized comparison of fluorouracil plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stage IIB-IVA carcinoma of the cervix with negative para-aortic lymph nodes: a Gynecologic Oncology Group and Southwest Oncology Group study. *J Clin Oncol* 1999; 17: 1339–1348.
64. Long HJ III, Bundy BN, Grendys EC Jr, et al. Randomized phase III trial of cisplatin with or without topotecan in carcinoma of the uterine cervix: a Gynecologic Oncology Group Study. *J Clin Oncol* 2005; 23: 4626–4633.
65. Omura GA, Blessing JA, Vaccarello L, et al. Randomized trial of cisplatin versus cisplatin plus mitolactol versus cisplatin plus ifosfamide in advanced squamous carcinoma of the cervix: a Gynecologic Oncology Group study. *J Clin Oncol* 1997; 15: 165–171.
66. Moore DH, Blessing JA, McQuellon RP, et al. Phase III study of cisplatin with or without paclitaxel in stage IVB, recurrent, or persistent squamous cell carcinoma of the cervix: a gynecologic oncology group study. *J Clin Oncol* 2004; 22: 3113–3119.
67. Thigpen T, Shingleton H, Homesley H, et al. Cisplatin in treatment of advanced or recurrent squamous cell carcinoma of the cervix: a phase II study of the Gynecologic Oncology Group. *Cancer* 1981; 48: 899–903.
68. Morris M, Brader KR, Levenback C, et al. Phase II study of vinorelbine in advanced and recurrent squamous cell carcinoma of the cervix. *J Clin Oncol* 1998; 16: 1094–1098.

69. McGuire WP, Blessing JA, Moore D, et al. Paclitaxel has moderate activity in squamous cervix cancer. A Gynecologic Oncology Group study. *J Clin Oncol* 1996; 14: 792–795.
70. Mudderspach LI, Blessing JA, Levenback C, Moore JL Jr. A phase II study of topotecan in patients with squamous cell carcinoma of the cervix: a gynecologic oncology group study. *Gynecol Oncol* 2001; 81: 213–215.
71. Bookman MA, Blessing JA, Hanjani P, et al. Topotecan in squamous cell carcinoma of the cervix: a phase II study of the Gynecologic Oncology Group. *Gynecol Oncol* 2000; 77: 446–449.
72. Rose PG, Blessing JA, Lele S, Abulafia O. Evaluation of pegylated liposomal doxorubicin (Doxil) as second-line chemotherapy of squamous cell carcinoma of the cervix: a phase II study of the Gynecologic Oncology Group. *Gynecol Oncol* 2006; 102: 210–213.
73. Sutton GP, Blessing JA, Adcock L, et al. Phase II study of ifosfamide and mesna in patients with previously-treated carcinoma of the cervix. A Gynecologic Oncology Group study. *Invest New Drugs* 1989; 7: 341–343.
74. Monk BJ, Sill M, McMeekin DS, et al. Phase III trial of four cisplatin-containing doublet combinations in stage IVB, recurrent or persistent cervical carcinoma: a Gynecologic Oncology Group study. *J Clin Oncol* 2009; 27: 4649–4655.
75. Monk BJ, Sill MW, Burger RA, et al. Phase II Trial of Bevacizumab in the Treatment of Persistent or Recurrent Squamous Cell Carcinoma of the Cervix: A Gynecologic Oncology Group Study. *J Clin Oncol* 2009; 27: 1069–1074.

Bloodless Surgery in Gynecologic Oncology

Nimesh P. Nagarsheth, MD,^{1,2} and Fahimeh Sasan, DO^{2,3}

¹Division of Gynecologic Oncology, Englewood Hospital and Medical Center, Englewood, NJ

²Mount Sinai School of Medicine, New York, NY

³Department of Obstetrics, Gynecology, and Reproductive Science, Englewood Hospital and Medical Center, Englewood, NJ

OUTLINE

RISKS OF TRANSFUSION

BLOODLESS SURGERY: PREOPERATIVE MANAGEMENT AND INTERVENTIONS

Phlebotomy

Intravenous Iron Treatment

(Preoperative Period)

Erythropoiesis-Stimulating Agents in the Preoperative Period

Autologous Blood Transfusion

INTRAOPERATIVE MANAGEMENT AND INTERVENTIONS

Intraoperative Autologous Transfusion

Acute Normovolemic Hemodilution

POSTOPERATIVE MANAGEMENT AND INTERVENTIONS

Case Example of Successful Execution of the Bloodless Surgery Protocol

Updates Regarding

Erythropoietin-Stimulating Agents

LATE-BREAKING STUDIES

CONCLUSION

ABSTRACT

Bloodless medicine and surgery is an evolving field in the practice of medicine designed to avoid allogeneic transfusions. Although this field has largely

developed in response to the growing needs of Jehovah's Witness patients refusing transfusions, all patients may potentially benefit from the avoidance of transfusions. The applications of bloodless techniques and strategies in the field of gynecologic oncology have been limited until recently, in part because of the generally large blood loss associated with gynecologic cancer surgeries. However, as advances in our understanding of bloodless surgery have developed and surgical techniques have been refined, the gynecologic oncology patient can now benefit from the avoidance of allogeneic transfusions. This review outlines bloodless surgery as it applies to gynecologic oncology patients and presents a structured approach to successfully managing these complex patients. *Mt Sinai J Med* 76:589–597, 2009. © 2009 Mount Sinai School of Medicine

Key Words: anemia, bloodless surgery, cancer, gynecology, Jehovah's Witness.

Bloodless medicine and surgery encompasses a broad spectrum of strategies and techniques designed to avoid allogeneic transfusions and minimize procedure-related blood loss.¹ Although measures to promote bloodless medicine and surgery have been traditionally performed to address patient populations that refuse transfusions (eg, Jehovah's Witness patients or patients in specific situations in which blood safety, supply, or compatibility is a concern), all patients can potentially benefit from the application of these strategies.² In this

All patients can potentially benefit from the application of bloodless medicine and surgery.

communication, we focus on bloodless medicine and surgery as it relates to the care of the surgical gynecology patient and the complete care of the gynecologic oncology patient.

Address Correspondence to:

Nimesh P. Nagarsheth

Division of Gynecologic
Oncology

Department of Obstetrics,
Gynecology, and Reproductive
Science

Mount Sinai Medical Center
New York, NY

Email: nimesh.nagarsheth@
gmail.com

The concepts of bloodless surgery can generally be divided into the 3 broad categories of preoperative, intraoperative, and postoperative interventions, all of which are designed to optimize the patient's health, minimize surgical blood loss, and avoid transfusions.¹ Importantly, in addition to the direct medical benefits of avoiding transfusions, bloodless management has been associated with a high level of patient satisfaction and overall improved patient care.⁴

RISKS OF TRANSFUSION

Transfusions have become a component of medical and surgical practices and have been used in clinical medicine for over 50 years.¹ However, despite longstanding beliefs regarding the benefits of blood product transfusion, there are a paucity of studies available to support existing clinical practice guidelines.⁴

Risks associated with blood transfusion include the transmission of infections, transfusion-related immunomodulation, acute lung injury, and hemolytic reactions.⁵ In addition to the fact that 25% of all banked blood is used in cancer surgery,⁶ the gynecologic oncology patient is particularly vulnerable to the risks of allogeneic transfusions. Blood product

Risks associated with blood transfusion include the transmission of infections, transfusion-related immunomodulation, acute lung injury, and hemolytic reactions.

transfusion involves risks of transmission of infectious agents, including but not limited to the bloodborne pathogens hepatitis B and C and human immunodeficiency virus. The American Association of Blood Banks reported the rates of transmission of these bloodborne pathogens for screened units of blood in 2007 as 1 in 137,000 for hepatitis B, fewer than 1 in 1,000,000 for hepatitis C, and fewer than 1 in 1,900,000 for human immunodeficiency virus.⁷

Transfusions have also been shown to increase the development of postoperative infection. In a recent retrospective analysis, the transfusion of packed red blood cells in an intensive care unit was associated with nosocomial infections, even after investigators adjusted for the probability of survival and age.⁸ Furthermore, blood transfusions have also been found to be independently associated

with a higher incidence of bacterial infections. In a prospective cohort study, the adjusted rate of bacterial infection was 4.8% for nontransfused patients, 15.2% for patients transfused with up to 2 units of red blood cells, 22.1% for patients transfused with 3 to 5 units of red blood cells, and 29% for patients transfused with 6 or more units of red blood cells ($P < 0.001$).⁹

Patients that receive multiple transfusions over time may experience allogeneic immunization, with subsequent transfusion reactions leading to limitations in the supply of compatible blood.¹⁰ In addition, induced immunosuppression results in a higher incidence of postoperative infections, reactivation of latent viruses,¹¹ and potential promotion of cancer growth.¹⁰ This phenomenon has been termed the immunomodulatory effect of blood transfusion.¹² Interestingly, a meta-analysis of the colorectal literature concluded that immunomodulation caused by perioperative blood transfusions is associated with an increased risk of recurrence of colorectal carcinoma and death from this malignancy.¹³

The leading cause of transfusion-associated mortality is transfusion-related acute lung injury (TRALI). The risk of TRALI is estimated to range from 1 in 1300 to 1 in 5000 per plasma-containing component transfusion.¹⁴ TRALI is a well-characterized clinical scenario that occurs during or within 6 hours of transfusion.¹⁵ The signs and symptoms include dyspnea, hypotension, and fever, which culminate in noncardiogenic pulmonary edema and respiratory compromise.^{15,16} Blood products that are associated with TRALI include whole blood, packed red blood cells, fresh frozen plasma, cryoprecipitate, and platelet concentrates. The etiology of TRALI may be attributable to anti-human leukocyte antigen or anti-granulocyte antibodies present in the plasma of multiparous females or donors who have received previous transfusions.¹⁷

Finally, an acute hemolytic transfusion reaction (AHTR) is a medical emergency that results from the rapid destruction of the donor's red blood and/or recipient's red blood. The risk of AHTR is 1 in 70,000 per unit. The etiology of AHTR includes ABO incompatibility or irregular alloantibodies induced by a former blood transfusion or gravidity on non-ABO antigens. A hemolytic transfusion reaction can present acutely with vague symptoms and subsequently lead to disseminated intravascular coagulation, renal failure, and hemorrhagic shock or can present as a delayed reaction 1 week after the transfusion, with jaundice, unexplained drops in hemoglobin, and fever of unknown origin. Fortunately, renal failure and death are uncommon in delayed reactions.¹⁸

BLOODLESS SURGERY: PREOPERATIVE MANAGEMENT AND INTERVENTIONS

Thorough preoperative preparation is essential to minimizing perioperative morbidity and mortality and begins with a comprehensive history and physical examination, with a special emphasis on a personal or family history of easy bruising or bleeding disorders.^{1,2} Optimization of preoperative hemoglobin levels with erythropoietin injections, iron, folate, or vitamin B12 supplementation should be performed when appropriate. Medications that may be involved in pharmacological coagulopathies (including acetylsalicylic acid and medications containing aspirin, nonsteroidal anti-inflammatory drugs, antiplatelet agents, and anticoagulants) should be discontinued prior to surgery. A preoperative hematology consult may be appropriate for the management of significant anemia or other complicated cases.² In addition, preoperative procedures that may aid in the minimization of surgical blood loss should be considered (Figure 1).¹ In particular, the gynecologic patient undergoing pelvic surgery may benefit from interventions such as uterine artery embolization.²

A critical component of bloodless medicine and surgery is appropriate and thorough discussion and counseling with patients with respect to their willingness to accept various components of blood (Figure 2).¹ We recently presented a review

A critical component of bloodless medicine and surgery is appropriate and thorough discussion and counseling with patients with respect to their willingness to accept various components of blood.

of patient responses to advanced directives signed by Jehovah's Witness patients prior to undergoing surgery in a gynecologic oncology service.¹⁹ Although our study confirmed that this population refuses to accept transfusions of whole blood, red cells, platelets, or plasma, we demonstrated that 98% of the patients agreed to accept a variety of other blood products, such as clotting factors, immunoglobulins, and cryoprecipitates. Our results underscore the importance of preoperative counseling in this patient population. Finally, the designation of a health care proxy and end-of-life decisions should be discussed with the patient prior to any surgical procedure.²

Phlebotomy

Another important consideration in preoperative assessment and planning is the restriction of diagnostic phlebotomy.² An article in the *New England Journal of Medicine* reviewed phlebotomy rates for diagnostic laboratory tests in adults and the resultant effects on transfusion requirements.²⁰ Of the 36 patients who received transfusions, 17 (47%) had large losses from phlebotomy (greater than 180 mL of red cells) that contributed to their transfusion requirements. The authors suggested approaches to reducing blood loss from phlebotomy, including the use of pediatric tubes for blood draws, the batching of requests for laboratory tests, and a review of the cumulative volume of blood removed from individual patients.²⁰

Intravenous Iron Treatment (Preoperative Period)

Preoperative intravenous (IV) iron administration may reduce both the proportion of patients requiring blood transfusions and the total amount of blood transfused, and high-dose bolus IV iron regimens are available if needed.²¹ A meta-analysis performed in 2008 by a panel of physicians from the Network for Advancement of Transfusion Alternatives studied the evidence on the efficacy and safety of IV iron administration to increase hemoglobin levels and reduce blood transfusions in patients undergoing surgery. On the basis of their findings, the authors recommended consideration of perioperative use of IV iron for orthopedic surgical patients and advocated further prospective trials to evaluate the role of IV iron use prior to other surgeries.²¹

Erythropoiesis-Stimulating Agents in the Preoperative Period

The preoperative use of epoetin alfa (Procrit) can help increase preoperative hemoglobin levels and may help avoid the need for a transfusion in the perioperative period.²²

Procrit label indications for use in the preoperative setting include 2 regimens: 600 U/kg weekly for 3 weeks prior to surgery or 300 U/kg daily for 10 days prior to surgery and for 4 days after surgery. Importantly, the 2 treatment regimens appear to be equally effective. There was no significant difference between the treatment groups in terms of the erythropoietic response, the mean number of units transfused per subject, or the mean hemoglobin levels in the postoperative period.²²

PREOPERATIVE ASSESSMENT**Counseling and Consents**

Date

Bloodless Protocol Patient Instruction Sheet		
Health Care Proxy Signed		
Surgical Consent		

Hemoglobin Optimization

Date

Iron, Folic Acid (vitamin B12)		
Erythropoietin Protocol		
Detection of Anemia:		
Hematology Consult (physician's discretion)		

Surgical Scheduling

Date

Cell Salvage Available		
Anesthesiologist Notification (preoperative assessment)		

INTRAOPERATIVE PLANNING**Surgical Equipment and Surgical Backup**

Date

Consideration of Acute Normovolemic Hemodilution		
Consideration Regarding Cell Salvage		
Hemostatics and Surgical Equipment Available		
Consideration of Surgical Backup		

OTHER

Date

Fig 1. Surgeon's checklist summary for bloodless surgery patients. This checklist is meant only to be an aid in the planning of an upcoming surgery. This is not an all-inclusive list, and not all of the listed options may apply to any particular patient.

Autologous Blood Transfusion

With respect to autologous blood donation, the utility, cost effectiveness, risks, and alternatives should be closely considered. In general, autologous blood donation is not acceptable to Jehovah's Witness patients as this patient population does not allow the transfusion of stored blood. In addition, 50% of patients undergoing preoperative autologous blood donation are anemic on the day of surgery.² In addition to a decrease in intracellular 2,3-diphosphoglycerate, which leads to a decrease in oxygen release, stored red blood cells undergo a number of other cellular changes, including biochemical, hormonal, inflammatory, and cellular structural changes, and they are prone to early destruction and have a shortened circulating half-life.⁵ Furthermore, the utility and cost effectiveness of preoperative autologous blood donation in a university-based gynecologic oncology practice were presented by Horowitz *et al.*²³ at the annual meeting of the Society of Gynecologic Oncologists in 2004. The authors demonstrated that preoperative

blood donation is expensive and does not guarantee that exposure to allogeneic blood will not occur (ie, patients may require additional units or may be given allogeneic blood because of clerical errors). They concluded that other methods of blood conservation, such as intraoperative hemodilution or preoperative erythropoietin and iron, may be safer and more cost-effective.

INTRAOPERATIVE MANAGEMENT AND INTERVENTIONS

Meticulous surgical technique remains the essential foundation of the intraoperative care of the bloodless surgery patient. Additional intraoperative

Meticulous surgical technique remains the essential foundation of the intraoperative care of the bloodless surgery patient.

I hereby consent to receive the blood products and/ or interventions marked below:

BLOOD PRODUCT/INTERVENTION	ACCEPT	REFUSE
Whole Blood		
Packed Red Blood Cells		
Platelets		
Fresh Frozen Plasma		
Cryoprecipitate		
Isolated Factor Preparations		
Albumin (blood fraction)		
Erythropoietin (contains albumin)		
Immune Globulins (blood fraction)		
Recombinant Isolated Factors (eg, NovoSeven made from the Human Factor VII gene in the laboratory with animal cells – no human serum or proteinis used)		
Human Topical Hemostatics [eg, human fibrin sealant may contain human thrombin and cryoprecipitate (Tisseal)]		
Nonhuman Topical Tissue Adhesives, Hemostatics [eg, absorbable collagen (Avitene), absorbable gelatin (Gelfoam), oxidized regenerated cellulose (Surgicel)]		
Intraoperative Hemodilution and Fractionation*		
Hemodialysis/Continuous Venovenous Hemofiltration		
Intraoperative Autologous Blood Salvage Cell Saver*		
Postoperative Blood Salvage/Reinfusion*		

*Extracorporeal circulation using a closed system without blood storage.

The above checklist correctly states my wishes regarding the use of blood products and interventions during my care at _____. The items checked in the refuse column are not to be given to me under any circumstances, even if physicians deem such necessary to preserve my life or health.

Patient's Name – Please Print

Witness's Name

Patient's Signature

Witness's Signature

Date(day/ month/ year)

Date(day/ month/ year)

Fig 2. Bloodless protocol patient instruction sheet. This has been adapted from the Bloodless Medicine and Surgery Program Checklist at Englewood Hospital and Medical Center. This checklist is provided solely as a guide and should not be considered a legal document. The interested reader should refer to local hospital policies and administration prior to implementing the use of this checklist or a similar one.

measures that should be considered include cell salvage (intraoperative autologous transfusion), acute normovolemic hemodilution, and the use of hemostatic agents (listed in Figure 1).^{24,25} The use of recombinant activated factor VII in cases of acute hemorrhage and coagulopathy may be considered in select cases.¹ The bloodless surgeon should stay current with advances in surgical technology, such as bipolar vessel sealing devices and ultrasonic energy devices, robotic surgery advances, and general advances in vaginal, laparoscopic, and traditional open surgery.

Intraoperative Autologous Transfusion

Intraoperative cell salvage is the process by which blood lost in the surgical field is collected, filtered, and washed to produce autologous red blood cells

for transfusion back to the patient.¹ However, cell salvage use during oncologic cases is controversial because of the risk of infusing cancer cells from the operative field into the vascular system. Six studies of patients who underwent operations for a variety of cancer cell types showed no change in the short-term or medium-term survival and no metastatic dissemination secondary to the use of intraoperative autotransfusion.²⁴ Importantly, the use of a leukocyte depletion filter to filter cell salvage blood during oncologic surgery has been strongly advocated in order to reduce the risk of intravenous reinfusion of cancer cells into the patient.^{1,24}

Importantly, in a randomized study of autologous and allogeneic blood transfusion–associated tumor recurrence in colorectal cancer surgery, allogeneic blood transfusions were found to be an independent risk factor for tumor recurrence.²⁵ In

addition, the irradiation of blood from intraoperative autologous transfusions in oncologic surgery is effective in preventing tumor cell dissemination.¹⁰

On the basis of these findings, we believe that it is reasonable to perform blood salvage with a leukocyte depletion filter in patients with gynecologic malignancies in whom allogeneic transfusions are not possible.

Acute Normovolemic Hemodilution

Acute normovolemic hemodilution is a safe and effective technique that can significantly reduce the need for allogeneic transfusions in surgical cases.²⁶ First, whole blood is collected from a patient via gravity into blood bags containing an anticoagulant before anticipated surgical blood loss (after the induction of anesthesia and prior to the commencement of surgery). Then, as blood volume is removed from the patient, IV crystalloid and/or colloid fluid is infused to maintain a normovolemic or euvoletic state. The amount of blood (red cell) loss that occurs during the diluted state is less than what would occur in a more concentrated state, and when the blood loss is completed, the whole blood that has been removed from the patient is reinfused intravenously.

A recent prospective randomized trial comparing acute normovolemic hemodilution to standard intraoperative management in patients undergoing major hepatic resection demonstrated a 50% reduction in the overall allogeneic red cell transfusion rate in patients undergoing hemodilution.²⁶ Patients undergoing acute normovolemic hemodilution were transfused less intraoperatively, had higher postoperative hemoglobin levels ($P = 0.01$), and overall required fewer red cell units (28 versus 47 units). In patients with intraoperative blood loss ≥ 800 mL, hemodilution reduced both the allogeneic red cell transfusion rate (18.2% versus 42.4%, $P = 0.045$) and the proportion of patients requiring fresh frozen plasma (21.1% versus 48.3%, $P = 0.025$). The application of acute normovolemic hemodilution in the setting of gynecologic oncology demonstrates that acute normovolemic hemodilution during radical hysterectomy does not significantly alter splanchnic or coagulation parameters.²⁷

POSTOPERATIVE MANAGEMENT AND INTERVENTIONS

Postoperative measures in bloodless surgery incorporate many of the techniques and treatments discussed

in the preoperative section of this article, including the limitation of phlebotomy, use of erythropoietin-stimulating agents (ESAs), iron therapy, and tolerance of postoperative anemia.¹ It is estimated that up to 90% of postoperative patients are anemic.²⁸ Although the main contributing factor of this anemia is perioperative blood loss, other processes such as inflammation-induced blunted erythropoiesis (often as a result of decreased iron availability) can confound the anemic picture.²⁹ In patients with significant blood loss anemia, IV iron alone can lead to a 5-fold erythropoietic response; however, some postoperative patients will benefit from ESAs in addition to IV iron.²²

Transfusion guidelines should not be based on "universal" hemoglobin levels that trigger a transfusion but should take into account a patient's individual ability to tolerate and compensate for the acute decrease in the hemoglobin concentration.^{30,31} Blood transfusions should be based on a patient's wishes as well as signs of inadequate tissue oxygenation that may occur at various hemoglobin concentrations depending on the patient's underlying diseases.^{32,33} As always, adequate IV fluid resuscitation should be employed to maintain adequate perfusion throughout the intraoperative and postoperative periods.

Furthermore, in anemic patients, tissue perfusion may be compromised not as a result of low hemoglobin levels but as a result of hypovolemia.³⁴ An important distinction is that the effects of blood transfusion on oxygen uptake are not as optimal as the effects of increasing blood flow because the rise in hematocrit increases the viscosity, which alters regional microvascular blood flow.³⁵

Finally, patients who were enrolled through the bloodless surgery program at a community hospital and underwent abdominal hysterectomy were significantly younger (average age, 43.4 versus 47.7 years) and incurred significantly lower hospital charges (average cost, \$8754 versus \$9539) than a non-bloodless comparison group. No significant difference between the group studied and the control group could be found in the average length of stay or the average change between preoperative and postoperative hemoglobin levels.³

Case Example of Successful Execution of the Bloodless Surgery Protocol

We recently reported the successful outcome of bloodless surgery in our gynecologic oncology service over an 8-year period.³⁶ Perhaps the most impressive use of bloodless surgery in our

experience was the case of a Jehovah's Witness patient with a 12.7-kg uterine leiomyosarcoma who underwent surgical resection.¹ The patient (who refused any blood transfusions) was enrolled in a bloodless surgical program and underwent surgical resection of a 40-cm pelvic mass. Several of the bloodless surgery techniques and interventions that have been described here were performed in this patient, including extensive preoperative counseling, preoperative uterine artery embolization, intraoperative cell salvage with a leukocyte depletion filter, factor VIIa administration, limited blood draws (with pediatric tubes), tolerance of anemia, and daily erythropoietin and iron administration. Although her intraoperative course was significant for severe anemia with a hemoglobin level of 2.5 g/dL and her postoperative course required a long-term intensive care unit stay, the patient fully regained her preoperative performance status.

Updates Regarding Erythropoietin-Stimulating Agents

The American Society of Clinical Oncology and American Society of Hematology recently reported recommendations for the use of epoetin.³⁴ For patients with chemotherapy-associated anemia, the committee continues to recommend initiating an ESA as hemoglobin approaches or falls below 10 g/dL for the purpose of increasing hemoglobin levels and decreasing transfusions. ESA treatment continues to be recommended for patients with low-risk myelodysplasia for similar reasons. There is no evidence showing increased survival as a result of ESA treatment. Conclusive evidence is lacking that initiating an ESA at hemoglobin levels greater than 10 g/dL either spares more patients from transfusion or substantially improves their quality of life.³⁴

Although many studies have demonstrated the efficacy of ESAs, a recent evaluation of randomized clinical trials comparing the use of ESAs to transfusion support alone in patients with active

cancer concluded that oncologic patients who received ESAs showed decreased survival, decreased duration of tumor control, and/or increased risk of thrombovascular events when the ESAs were administered to achieve hemoglobin levels of 12 g/dL in patients with nonmyeloid malignancies.³⁵

In July 2008, the US Food and Drug Administration Oncologic Drugs Advisory Committee recommended limits on the use of synthetic erythropoietin products to treat anemia in many cancer patients. Table 1 details the new Food and Drug Administration recommendations and changes in product labeling.³⁷ In general, the decision to use erythropoietin products in our patients is made on a case-by-case basis.

LATE-BREAKING STUDIES

Recently, a case series was published documenting the long-term outcomes of 3 patients with gross intra-abdominal malignancies that underwent surgical resection in a gynecologic oncology bloodless surgery service. All 3 patients underwent reinfusion of cell salvage blood (after filtration with a leukocyte depletion filter) during their procedures. Two of the 3 cases did not show any evidence of metastases over a significant follow-up period. One patient did have postoperative evidence of hematogenous dissemination; however, this patient had known liver metastases at the time of her initial diagnosis and therefore had evidence of hematogenous dissemination prior to her index surgical procedure. This study adds further credibility to the recommendation that blood salvage with filtration using a leukocyte depletion filter may be acceptable in patients with gross intra-abdominal gynecologic malignancies in whom allogeneic transfusion is not possible. Given the potential risks of reinfusion of cell salvage blood from a surgical field contaminated with cancer cells, a large prospective study investigating this important issue is not feasible. Therefore, we must continue to accumulate retrospective data through case reports and case

Table 1. *Epoen (Epoetin Alfa): The Complete Response and Safety Labeling Change Order of the Food and Drug Administration (July 30, 2008).*

1. ESAs shortened overall survival and/or increased the risk of tumor progression or recurrence in some clinical studies in patients with breast, non-small cell lung, head and neck, lymphoid, and cervical cancers.
2. To decrease these risks as well as the risk of serious cardiac and thrombovascular events, use the lowest dose needed to avoid red blood cell transfusions.
3. Use ESAs only for the treatment of anemia due to concomitant myelosuppressive chemotherapy.
4. ESAs are not indicated for patients receiving myelosuppressive therapy when the anticipated outcome is cure.
5. Discontinue ESAs after the completion of a chemotherapy course.

NOTE: This table was adapted from the US Food and Drug Administration Web site.³⁷

Abbreviation: ESA, erythropoietin-stimulating agent.

series such as this one to make a reasonable assessment of the risks and benefits of this intervention.³⁸

CONCLUSION

In conclusion, bloodless medicine and surgical techniques include optimization of the preoperative hematocrit, meticulous management of intraoperative blood loss, utilization of nonallogeneic blood products, iron, and ESAs, and postoperative tolerance of anemia. These techniques are feasible in all gynecologic surgical cases, and utilization of these methods can ultimately help reduce blood transfusion rates.¹

DISCLOSURES

Potential conflict of interest: Dr. Nagarsheth is a consultant for Covidien, Inc.

REFERENCES

- Nagarsheth NP, Shander A, Malovany R, et al. Bloodless surgery in a Jehovah's Witness patient with a 12.7-kg uterine leiomyosarcoma. *J Surg Educ* 2007; 64: 212–219.
- Goodnough LT, Shander A, Spence R. Bloodless medicine: clinical care without allogeneic blood transfusion. *Transfusion* 2003; 43: 668–676.
- DeCastro R. Bloodless surgery: establishment of a program for the special medical needs of the Jehovah's Witness community—the gynecologic surgery experience at a community hospital. *Am J Obstet Gynecol* 1999; 180: 1491–1498.
- Hardy JF. Current status of transfusion triggers for red blood cell concentrates. *Transfus Apher Sci* 2004; 31: 55–66.
- Hajjar LA, Auler Junior JO, Santos L, Galas F. Blood transfusion in critically ill patients: state of the art. *Clinics* 2007; 62: 507–524.
- Friedman BA, Burns TL, Schork MA. A study of blood utilization by diagnosis, month of transfusion, and geographic region of the United States. *Transfusion* 1979; 19: 511–525.
- American Association of Blood Banks. Transfusion-transmitted diseases, April 2007. Available at: <http://www.aabb.org>. Accessed September 2009.
- Taylor RW, Manganaro L, O'Brien J, et al. Impact of allogeneic packed red blood cell transfusion on nosocomial infection rates in the critically ill patient. *Crit Care Med* 2002; 30: 2249–2254.
- Chelemer SB, Prato BS, Cox PM, et al. Association of bacterial infection and red blood cell transfusion after coronary artery bypass surgery. *Ann Thorac Surg* 2002; 73: 138–142.
- Hansen E, Kneuchel R, Altmeppen J, Taeger K. Blood irradiation for intraoperative autotransfusion in cancer surgery: demonstration of efficient elimination of contaminating tumor cells. *Transfusion* 1999; 39: 608–615.
- Blumberg N. Allogeneic transfusion and infection: economic and clinical implications. *Semin Hematol* 1997; 34: 34–40.
- Klein H. Immunomodulatory aspects of transfusion: a once and future risk? *Anesthesiology* 1999; 91: 861–865.
- McAlister F, Clark H, Wells P, Laupacis A. Perioperative allogeneic blood transfusion does not cause adverse sequelae in patients with cancer: a meta-analysis of unconfounded studies. *Br J Surg* 1998; 85: 171–178.
- Shander A, Popovsky MA. Understanding the consequences of transfusion-related acute lung injury. *Chest* 2005; 128: 598s–604s.
- Toy P, Popovsky MA, Abraham E, et al. Transfusion-related acute lung injury: definition and review. *Crit Care Med* 2005; 33: 721–726.
- Sachs UJ. The pathogenesis of transfusion-related acute lung injury and how to avoid this serious adverse reaction of transfusion. *Transfus Apher Sci* 2007; 37: 273–282.
- Popovsky MA, Chaplan HC, Moore SB. Transfusion-related lung injury: a neglected serious complication of hemotherapy. *Transfusion* 1992; 32: 589–592.
- Popovsky MA. Transfusion-related acute lung injury: incidence, pathogenesis and the role of multicomponent apheresis in its prevention. *Transfus Med Hemother* 2008; 35: 76–79.
- Nagarsheth NP, Gupta N, Moshier E, et al. Responses on advance directives by Jehovah's Witnesses on a gynecologic oncology service. Abstracts presented for the 39th annual meeting of the Society of Gynecologic Oncologists. *Gynecol Oncol* 2007; 108: s1–s182.
- Smoller BR, Kruskall MS. Phlebotomy for diagnostic laboratory tests in adults: pattern of use and effect on transfusion requirements. *N Engl J Med* 1986; 314: 1233–1235.
- Beris P, Munoz M, Garcia-Erce A, et al. Perioperative anaemia management: consensus statement on the role of intravenous iron. *Br J Anaesth* 2008; 100: 599–604.
- Procrit (epoetin alfa) prescribing information. Available at: <http://www.procrit.com/procrit/shared/OBI/PI/ProcritBooklet.pdf#page=1>. Accessed September 2009.
- Horowitz NS, Gibb RK, Menegakis NE, et al. Utility and cost effectiveness of preoperative autologous blood donation in a university-based gynecologic oncology practice. *Gynecol Oncol* 2004; 92: 1–512.
- Élias D, Valérie B, Valérie L. Use of the cell saver in oncologic surgery. Transfusion alternatives. *Transfus Med* 2008; 3: 25–28.
- Heiss MM, Mempel W, Jauch KW, et al. Beneficial effects of autologous blood transfusion on infectious complications after colorectal cancer surgery. *Lancet* 1993; 342: 1328–1333.
- Jarnagin WR. A prospective randomized trial of acute normovolemic hemodilution compared to standard intraoperative management in patients undergoing major hepatic resection. *Ann Surg* 2008; 248: 360–369.
- Santoso JT, Hannigan EV, Levine L, et al. Effect of hemodilution on tissue perfusion and blood coagulation during radical hysterectomy. *Gynecol Oncol* 2001; 82: 252–256.
- Shander A, Goodnough L. Objectives and limitations of bloodless medical care. *Curr Opin Hematol* 2006; 13: 462–470.

29. Munoz M, Breymann C, Garcia-Erce JA, et al. Efficacy and safety of intravenous iron therapy as an alternative/adjunct to allogeneic blood. *Vox Sang* 2008; 94: 172–183.
30. Madjdpour C, Donat S, Richard B. Anemia and perioperative red blood cell transfusion: a matter of tolerance. *Crit Care Med* 2006; 34: s102–s108.
31. Shander A. Surgery without blood. *Crit Care Med* 2003; 31: s709–s714.
32. Stovall TG. Clinical experience with epoetin alfa in the management of hemoglobin levels in orthopedic surgery and cancer: implications for use in gynecologic surgery. *J Reprod Med* 2001; 46: 531–538.
33. Suzuki Y, Tokuda Y, Fujiwara Y, et al. Weekly epoetin beta maintains haemoglobin levels and improves quality of life in patients with non-myeloid malignancies receiving chemotherapy. *Jpn J Clin Oncol* 2008; 38: 214–221.
34. Rizzo JD, Somerfield MR, Hagerty KL, et al. Use of epoetin and darbopoetin in patients with cancer: 2007 American Society of Clinical Oncology/American Society of Hematology clinical practice guideline update. *J Clin Oncol* 2008; 26: 132–149.
35. Juneja V, Keegan P, Gootenberg JE, et al. Continuing reassessment of the risks of erythropoiesis-stimulating agents in patients with cancer. *Clin Cancer Res* 2008; 14: 3242–3247.
36. Nagarsheth NP, Gupta A, Gretz HF, et al. Feasibility of bloodless surgery on a gynecologic oncology service. *Int J Gynecol Cancer* 2006; 16: 624–694.
37. Epogen (epoetin alfa): FDA complete response and safety labeling change order: July 30, 2008. Available at: http://www.accessdata.fda.gov/drugsatfda_docs/label/2008/103234s5195sPI.pdf. Accessed September 2009.
38. Nagarsheth, NP, Sharma T, Shander A, Awan A. Blood salvage use in gynecologic oncology. *Transfusion*. doi://10.1111/j.1537-2995.2009.02256.x. Available at: <http://www3.interscience.wiley.com>.

SPECIAL FEATURE

Breast Magnetic Resonance Imaging: An Overview for Nonradiologists

Laurie Margolies, MD, FACR, and Richard Ha, MD

Mount Sinai School of Medicine, New York, NY

OUTLINE

CLINICAL APPLICATIONS OF BREAST MAGNETIC
RESONANCE IMAGING

**Additional Evaluation of
Mammographic, Ultrasound, or
Palpable Abnormalities
Screening Women at High Risk for
Breast Cancer
Evaluation of the Extent of Disease
Other Indications**

ORDERING BREAST RESONANCE IMAGING

BREAST MAGNETIC RESONANCE IMAGING REPORT

CONCLUSION

ABSTRACT

Magnetic resonance imaging is a major component of breast imaging. Many studies have shown that magnetic resonance imaging is the most sensitive imaging method for detecting invasive breast cancer in comparison with mammography, ultrasound, and clinical breast examinations. Evidence-based clinical indications for breast magnetic resonance imaging include screening patients at high risk for breast cancer, including those with breast/ovarian cancer genes (BRCA1 and BRCA2), those who are untested first-degree relatives of carriers of these genes, those whose lifetime risk of developing breast cancer is 20% to 25% or greater, those who had chest

radiation when they were 10 to 30 years old, and those who have or are first-degree relatives of people with Li-Fraumeni syndrome, Cowden syndrome, or Bannayan-Riley-Ruvalcaba syndrome. Breast magnetic resonance imaging is performed in conjunction with mammography and does not replace mammography. Outside of the screening population, utilization of breast magnetic resonance imaging for newly diagnosed breast cancer patients and its use as a problem-solving technique for equivocal mammographic or clinical findings remain controversial. An understanding of the current evidence facilitates appropriate utilization of this important medical resource. This article discusses indications for ordering breast magnetic resonance imaging and how to read the breast magnetic resonance imaging report and understand the lexicon used. *Mt Sinai J Med* 76:598–605, 2009. © 2009 Mount Sinai School of Medicine

Key Words: breast cancer screening, breast magnetic resonance imaging indications, magnetic resonance imaging.

Breast magnetic resonance imaging (MRI) examinations compose a small percentage of the MRI examinations performed in the United States.^{1,2} It is, however, a growing percentage. The overall volume of MRI studies has grown only 3% in the past 4 years, but breast MRI volume has doubled since 2003. Breast MRI was performed 314,000 times in 2003; this rose to 645,000 in 2007.^{1,2} Almost all breast MRI examinations performed in the United States today are performed at 1.5 or 3 T; most weaker magnets are not considered suitable for breast MRI. Three-tesla magnets provide increased resolution and speed, although equivalent images can be obtained at 1.5 T with breast coils that are now often 8 or 16 channels; until recently, many breast coils were only 4 channels.²

Radiologists, patients, breast surgeons, oncologists, gynecologists, and primary care physicians

Address Correspondence to:

Laurie Margolies, MD, FACR
Mount Sinai School of Medicine
One Gustave Levy Place,
Box 1234
Email: laurie.margolies@
mountsinai.org

are all involved in the rapid growth of breast MRI. Evidence-based indications have been established and are growing, whereas other indications remain controversial and under continued investigation. The sensitivity of breast MRI has been proven by many studies; its specificity, benefit to survival, and overall effect on patient care—timeliness of care, mastectomy rates, and associated anxiety—have caused some to question its current and expanding role in patient care.

The sensitivity of breast MRI has been proven by many studies; its specificity, benefit to survival, and overall effect on patient care—timeliness of care, mastectomy rates, and associated anxiety—have caused some to question its current and expanding role in patient care.

CLINICAL APPLICATIONS OF BREAST MAGNETIC RESONANCE IMAGING

Additional Evaluation of Mammographic, Ultrasound, or Palpable Abnormalities

Clinical studies of breast MRI have investigated its potential role in further evaluating an abnormal mammographic, ultrasound, or clinical breast examination finding. It was hoped that MRI could reduce the number of benign biopsies. If an MRI examination were normal, it was hypothesized that biopsy would not be needed. However, a study by the International Breast MRI Consortium in 2004 demonstrated that, despite the high sensitivity of MRI, 12% of cancers identified by mammography or clinical breast examination were not detected by MRI.³ Breast MRI does not detect all cases of ductal carcinoma in situ or all small carcinomas (<3 mm). According to Friedrich,⁴ 5% to 12% of invasive carcinomas are not recognized during MRI because they lack the typical criteria for carcinomas. This is likely because neoangiogenesis may not be present or may be too minimal to produce kinetic abnormalities. The morphologic abnormalities of a typical cancer may be missing. Mammography finds ductal carcinoma in situ and some small cancers that MRI does not. MRI should not, therefore, be used to overrule a mammographic, ultrasound, or clinical recommendation for biopsy;

that is, a negative MRI report should not dissuade one from biopsying a suspicious finding. This must often be explained to patients who may regard a normal MRI report as a definitive “no cancer is present” examination.

A study by the International Breast MRI Consortium in 2004 demonstrated that, despite the high sensitivity of MRI, 12% of cancers identified by mammography or clinical breast examination were not detected by MRI. Mammography finds ductal carcinoma in situ and some small cancers that MRI does not. MRI should not, therefore, be used to overrule a mammographic, ultrasound, or clinical recommendation for biopsy.

Screening Women at High Risk for Breast Cancer

MRI is not an appropriate screening tool for the entire population for many reasons, including its high cost, limited MRI resources, the length of the examination, the risk of intravenous contrast, and the relatively high false positive rate and associated lack of specificity. However, it has been shown to be useful in very select populations. In 2007, the American Cancer Society published its recommendations concerning breast MRI screening. They remain current, are widely used, and are often quoted.⁵ The panel found that there was sufficient evidence to recommend yearly breast MRI in addition to yearly mammography for those patients who have a BRCA1 or BRCA2 mutation and for those patients who have a first-degree relative with the mutation but have not been tested. Others deemed to benefit from yearly breast MRI include those who have a 20% to 25% lifetime risk of developing breast cancer based on a risk prediction model. These models are in evolution, and risk results vary, depending on which model is chosen. At <http://www.cancer.gov>, one can find an interactive tool for measuring the risk of invasive breast cancer in those patients over the age of 35. Women who were treated with chest radiation between the ages of 10 and 30 (ie, those who

received mantle field radiation for Hodgkin's disease) or had some rare syndromes were found to have a proven benefit from breast MRI. These syndromes include Li-Fraumeni syndrome, Cowden syndrome, and Bannayan-Riley-Ruvalcaba syndrome.⁵

Evaluation of the Extent of Disease

The American Cancer Society guidelines deem MRI to be of uncertain value for breast cancer survivors, but numerous studies have addressed the role of breast MRI in the patient newly diagnosed with breast cancer. A study by Lehman *et al.*⁶ showed that MRI can be useful for finding tumors in the contralateral breast of women newly diagnosed with the disease. They studied 969 newly diagnosed breast cancer patients to see if MRI could find contralateral cancers that mammograms and physical examinations

The American Cancer Society guidelines deem MRI to be of uncertain value for breast cancer survivors, but numerous studies have addressed the role of breast MRI in the patient newly diagnosed with breast cancer. Utilization of MRI to detect additional cancers earlier could influence treatment decisions and might reduce or eliminate additional surgery and chemotherapy or conversely could lead to unnecessary interventions and delays in treatment.

missed. MRI found 30 early-stage tumors that the other tests did not detect and missed only 3. Previously, Harms *et al.*⁷ demonstrated that MRI could identify otherwise occult multicentric and multifocal disease in women with breast cancer. They compared magnetic resonance images with mastectomy specimens. MRI detected additional disease in 37% of specimens. Studies by Bedrosian *et al.*⁸ and Tan *et al.*⁹ found that MRI changed the management of 23% of breast cancer patients scheduled for breast-conserving therapy. Recently, Lee¹⁰ presented a study at the American Roentgen Ray Society; she studied 110 patients who had presurgical breast MRI. When only 14 days were

added to the time from diagnosis to surgery, 16 additional cancers were found in 13 women. MRI changed the treatment in 28% of women: some had more extensive surgery, whereas others opted for mastectomy with or without additional biopsy proof of disease. Utilization of MRI to detect additional cancers earlier could influence treatment decisions and might reduce or eliminate additional surgery and chemotherapy or conversely could lead to unnecessary interventions and delays in treatment.

Other Indications

Currently, other accepted indications for breast MRI are monitoring patients undergoing neoadjuvant chemotherapy and evaluating patients with axillary adenocarcinoma with an unknown primary site. Breast MRI is useful for the evaluation of breast implant integrity.¹¹ The evaluation of breast implant integrity is performed without intravenous contrast administration and is not used for cancer detection.

ORDERING BREAST RESONANCE IMAGING

Before considering breast MRI for an individual patient, one must be certain that she is a suitable candidate for the examination. In addition to meeting all the general MRI safety requirements, such as not having a pacemaker, the patient must be able to tolerate the 30- to 45-minute examination. She must be able to remain prone, usually with her arms extended over her head. Some facilities have breast coils that allow patients to be scanned with their arms by their side. The patient's glomerular filtration rate/renal function must allow her to have intravenous contrast. Most MRI units have a 300- or 350-pound weight limit. Claustrophobia is an issue for some patients. Even after the decision is made, one may still face hurdles from an insurance carrier or its precertification division.

Because the breast is a hormonally active organ, it is best to schedule a breast MRI examination during days 7 to 14 of the menstrual cycle or have a patient discontinue hormonal therapy for 2 months prior to the examination. This limits background enhancement. However, one should not delay surgery while waiting for the best time to perform a breast MRI examination. If background enhancement is a problem, the examination can always be repeated.

Bilateral breast MRI is strongly preferred to unilateral breast MRI. Recent technological advances

have allowed examinations to have excellent spatial and temporal resolution simultaneously. Patients no longer need to have one breast scanned and then return to have the other scanned days later. Benefits of bilateral examinations include the ability to compare background enhancement. The breasts should have equivalent backgrounds; if they do not, this must be explained. Scanning the opposite breast for occult disease and the chest wall for occult recurrence are also valid reasons for almost always doing bilateral breast MRI.

BREAST MAGNETIC RESONANCE IMAGING REPORT

After one orders breast MRI and the patient has the examination, a report is generated. The American College of Radiology developed the Breast Imaging Reporting and Data System, which includes an MRI lexicon that is recommended.¹² The standardized

After one orders breast MRI and the patient has the examination, a report is generated. The American College of Radiology developed the Breast Imaging Reporting and Data System, which includes an MRI lexicon that is recommended.

terminology, assessment categories, and recommendation categories facilitate communication across institutions and are useful even when protocols and interpretation strategies differ.

The breast MRI report may contain many features but should include the relevant breast history: indications for the study and any prior breast procedures. Prior studies are requested for comparison whenever possible. These might include prior MRI examinations, ultrasound examinations, and mammograms. The technique, including the amount of intravenous contrast, is included in the report. Including the date of the last menstrual period is often helpful in putting findings in perspective.

By the inclusion of the degree of background enhancement, the reader is given an idea of the examination's sensitivity. The greater the background enhancement, the lower the sensitivity. When a finding is present, the report may include its location within the breast, location on a specific image number, size, and enhancement pattern. The examination almost always visualizes a portion of the spine and liver, and comments are often made about them. Breast MRI examinations have found occult hepatic and renal masses. Finally, the report should give an impression and recommendations for further evaluation or follow-up.

Background enhancement is generally defined as minimal when there is little (less than 10%) background enhancement and as marked when there is extensive (greater than 75%) background enhancement that could limit sensitivity. Mild enhancement means that there is 10% to 50% background enhancement, and moderate background enhancement is 50% to 75% (Figures 1 and 2).

Lesions are defined as masses or nonmasses and are evaluated on images with different pulse sequences: T1 and T2 weighting (or equivalent images) and precontrast and postcontrast images

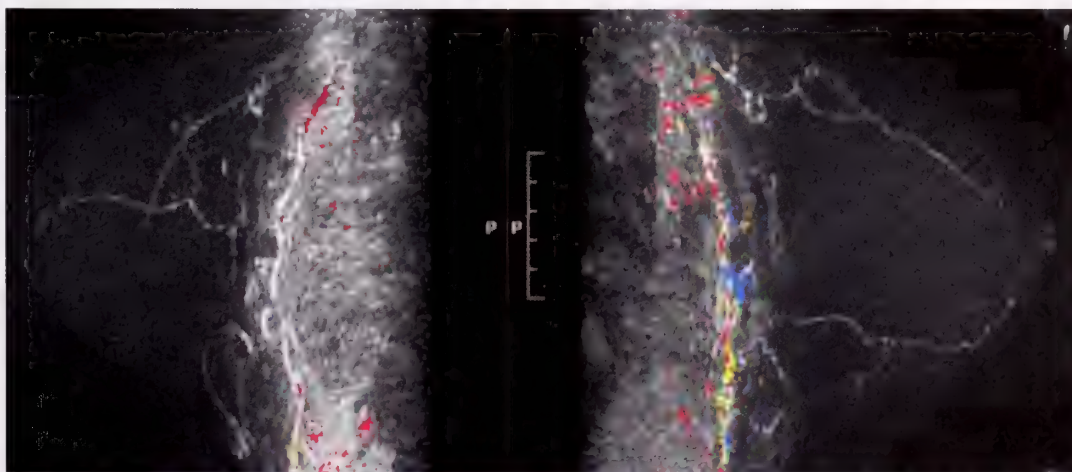


Fig 1. Breast magnetic resonance imaging: minimal background enhancement.

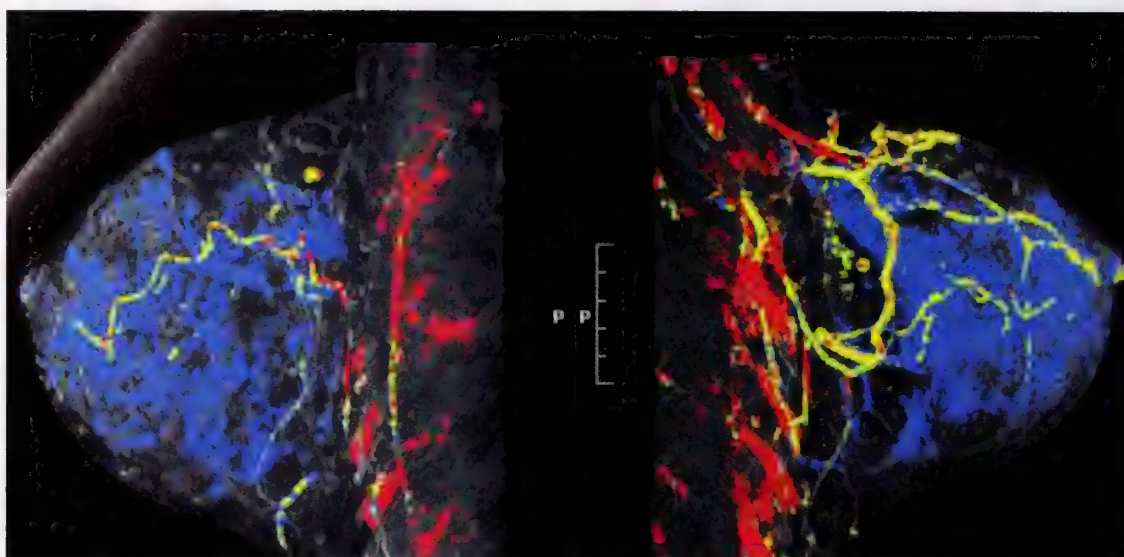


Fig 2. Breast magnetic resonance imaging: moderate background enhancement.

are used to facilitate diagnosis. Masses have 3-dimensional shapes and are generally larger than 5 mm. Their description will likely include the location, shape, and margins as well as enhancement kinetics. The shape can be round, oval, lobular, or irregular, and the margins can be smooth, irregular, or spiculated. The larger the lesion is, the easier it is to characterize the margins, and the higher the spatial resolution is, the easier it is to characterize the margins. If the spatial resolution is not sufficient, irregular margins can appear smooth. Tiny lesions can appear to have smooth margins when in fact they are irregular.

Figures 3 and 4 show the classic appearance of simple cysts on MRI: well-defined, low-signal masses on T1-weighted sequences that have a high signal on T2 sequences (Figure 3) and a signal void on subtracted images from a dynamic contrast-enhanced series (Figure 4).

Foci are usually round areas of enhancement that are less than 5 mm. In general, a mass must prove itself to be benign, and a focus must prove itself to be worrisome. Most foci are safely ignored.¹³

Kinetics are evaluated for each mass. Usually, the breast MRI examination includes a dynamic fat suppressed sequence with a precontrast image and 1 to 4 postcontrast images acquired over about 500 seconds. Masses can have no enhancement, homogeneous or heterogeneous enhancement, or rim enhancement and may have internal enhancing or nonenhancing septations. One analyzes the curve for the speed of the initial rise (ie, the change between the precontrast run and the first dynamic run) and for the change in the enhancement over

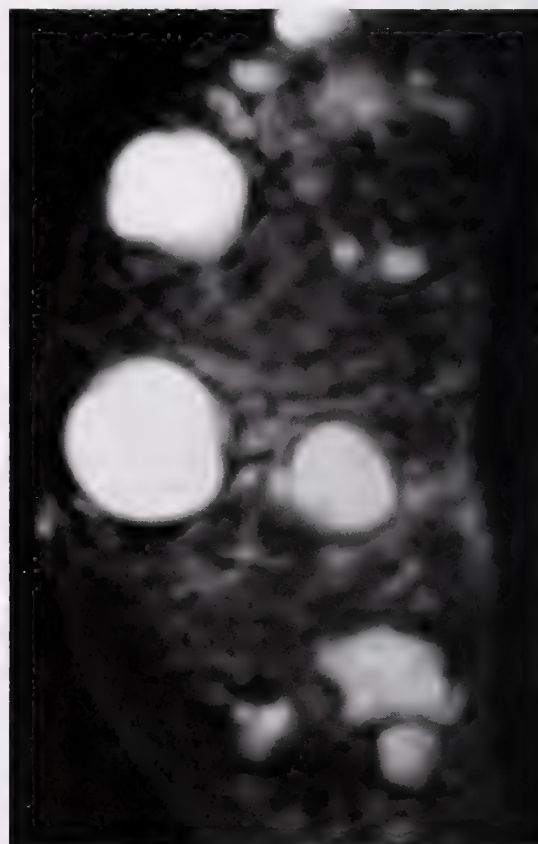


Fig 3. Breast magnetic resonance imaging: classic appearance of simple cysts.

time. Subtracted images are used to facilitate interpretation. The precontrast image is subtracted from the contrast-enhanced images. In general, malignant lesions will show heterogeneous enhancement with

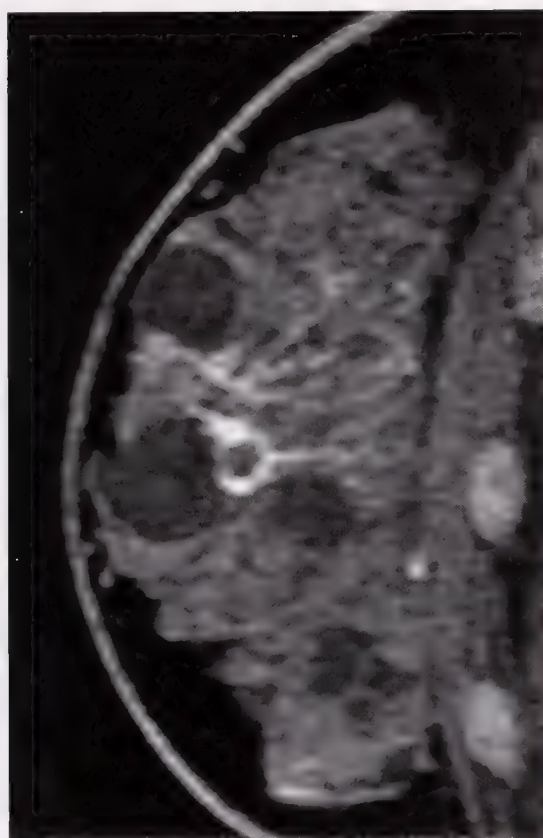


Fig 4. Breast magnetic resonance imaging: signal void present on a subtracted image, thus proving no enhancement within a mass.

some rapid washout (Figure 5A,B). Delayed images can also show plateau enhancement (no significant change during the dynamic scans; Figure 6) or persistent enhancement (progressive enhancement over time; Figure 7). Persistently increasing enhancement generally indicates a benign curve, but there is considerable overlap; benign lesions can show washout, and malignant lesions can have little or no washout. The classic benign lymph node, for example, has extensive areas of rapid washout.

Non-mass-like enhancement is an important term that is often seen in an MRI report. Stippled enhancement indicates normal or fibrocystic breast tissue, and looks like many foci or punctate dots that are diffusely distributed. Ductal carcinoma in situ is often seen on MRI as an area of clumped non-mass-like enhancement that resembles cobblestones. It may be linear and may point to the nipple; it can occupy a quadrant or segment of the breast and can be bilateral. Figure 8 shows a segmental area of clumped enhancement.

When a worrisome mass is depicted on MRI, it can be biopsied under MRI guidance. If, however,

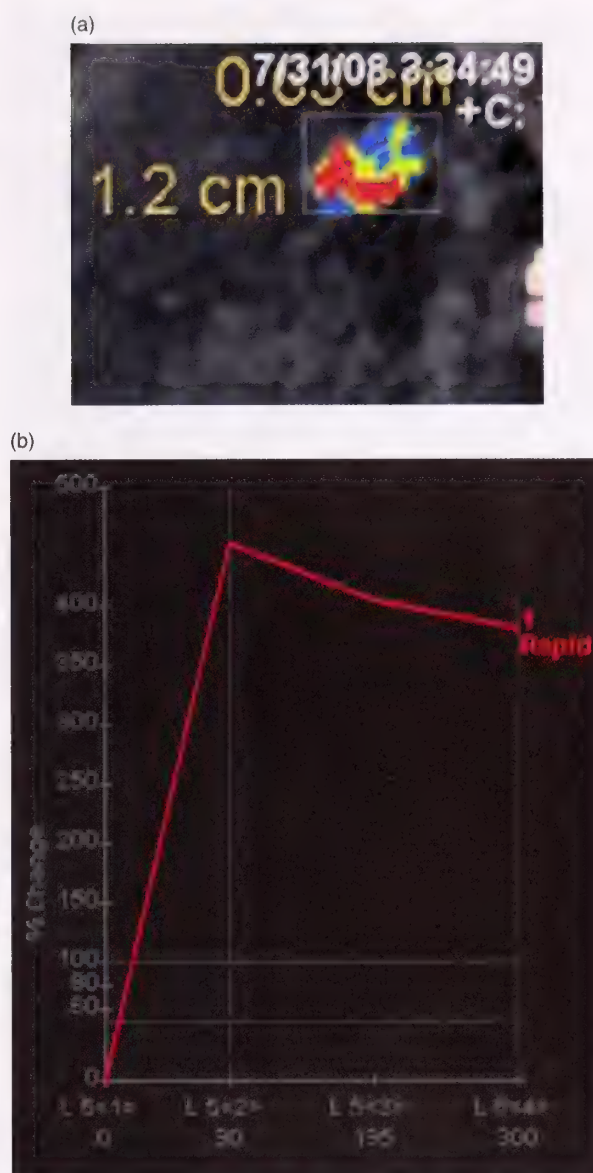


Fig 5. Breast magnetic resonance imaging: malignant lesions showing heterogeneous enhancement with some rapid washout.

it can be seen on ultrasound, it is much easier for the patient to have ultrasound-guided biopsy rather than MRI-guided biopsy. This is the rationale for "second look" ultrasound.¹⁴ It is imperative, of course, to be certain that any lesion evaluated by ultrasound with the patient in the supine position is the same finding that was seen during the prone MRI examination. The clumped enhancement of ductal carcinoma in situ often has no ultrasound or mammographic correlate and is thus limited to MRI-guided biopsy. DeMartini *et al.*¹⁵ found an ultrasound correlate for 58% of masses, 30% of areas of non-mass-like enhancement, and 37% of foci. A higher positive biopsy rate was found for

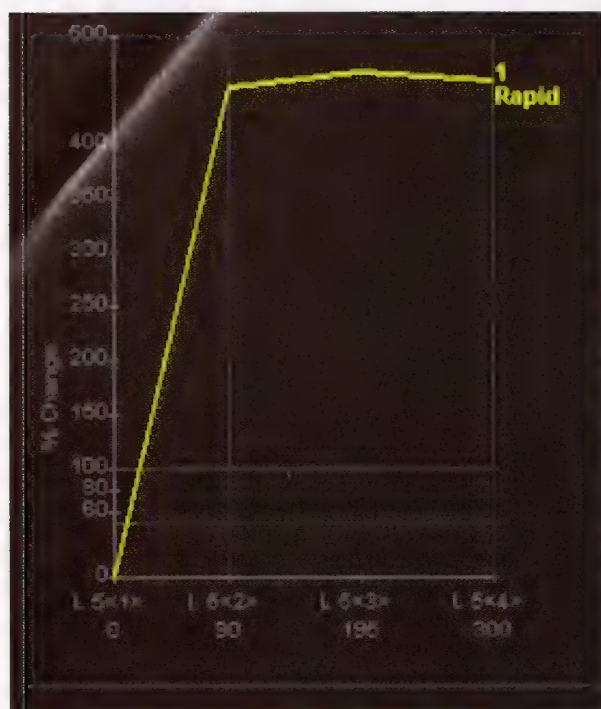


Fig 6. Breast magnetic resonance imaging: delayed image showing plateau enhancement (ie, no significant change during dynamic scans).

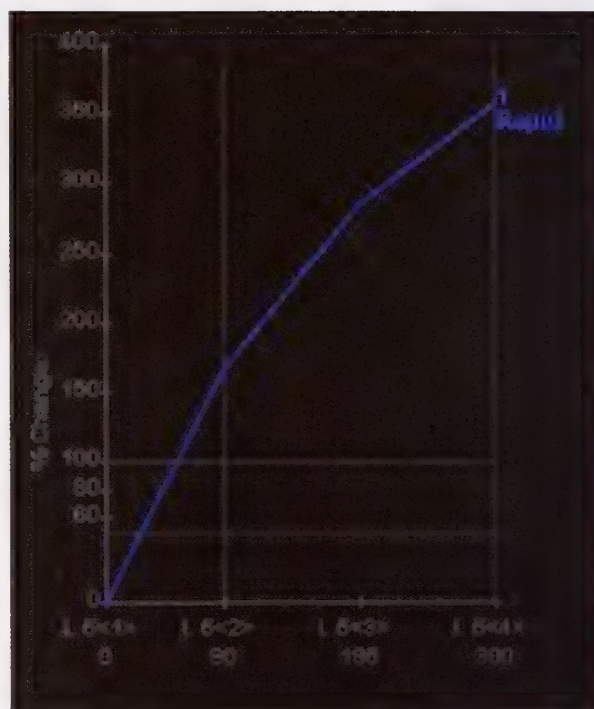


Fig 7. Breast magnetic resonance imaging: delayed image showing persistent enhancement (ie, progressive enhancement over time).

CONCLUSION

those lesions that were seen on ultrasound versus those that were sonographically occult: 36% versus 22%.¹⁵

The growing field of breast MRI is challenging not only to breast imagers but also to payers who are deciding what to cover, patients who must decide

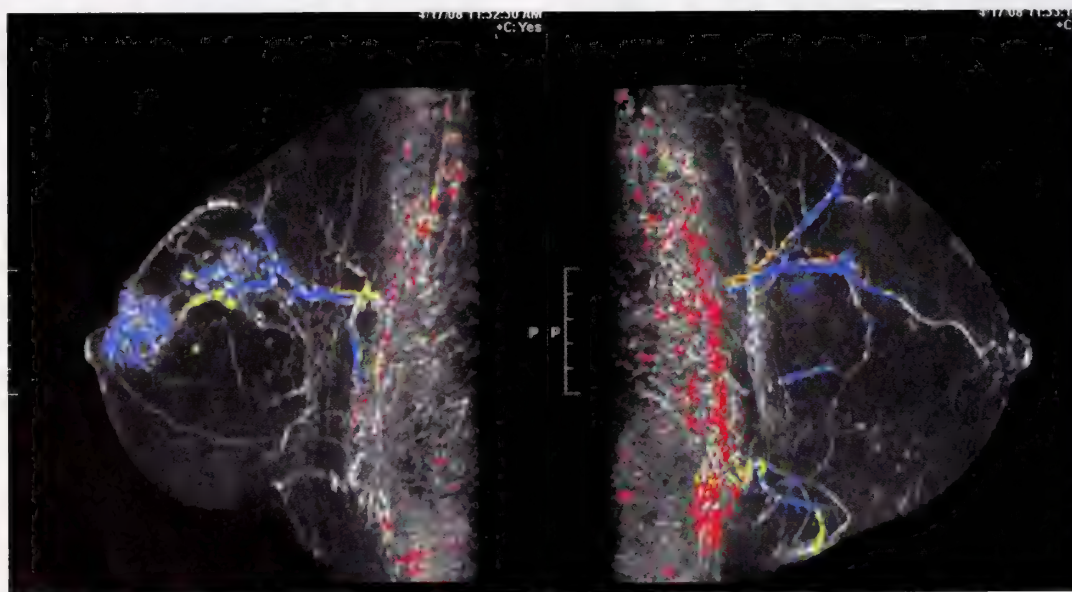


Fig 8. Breast magnetic resonance imaging: ductal carcinoma in situ showing a segmental area of clumped enhancement.

about having the examination, patients who must accept the positioning and anxiety inherent in the examination, and ordering physicians who navigate this burgeoning field. Researchers continue to define and debate the proper use of breast MRI in the evaluation of high-risk patients and those with breast cancer. As experience with breast MRI grows, it is likely that the examination times will be shorter and that the examinations will be more sensitive and specific.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Health Imaging News. June 6, 2008. Available at: <http://www.healthimaging.com>. Accessed September 2009.
2. Bassett LW. National trends and practices in breast MRI. *AJR Am J Roentgenol* 2008; 191: 332–339.
3. Schnall MD, Blume J, Bluemke D, et al., for the International Breast MRI Consortium. MRI detection of multi focal breast carcinoma: report from the International Breast MRI Consortium. *J Clin Oncol* 2004; 22(suppl): 14S.
4. Friedrich M. MRI of the breast: state of the art. *Eur Radiol* 1998; 8: 707–725.
5. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007; 57: 75–89.
6. Lehman CD, Gatsonis C, Kuhl CK, et al., for the ACRIN Trial 6667 Investigators Group. MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. *N Engl J Med* 2007; 356: 1295–1303.
7. Harms SE, Flamig DP, Hesley KL, et al. MR imaging of the breast with rotating delivery of excitation off resonance: clinical experience with pathologic correlation. *Radiology* 1993; 187: 493–501.
8. Bedrosian I, Mick R, Orel SG, et al. Changes in the surgical management of patients with breast carcinoma based on preoperative magnetic resonance imaging. *Cancer* 2003; 98: 468–473.
9. Tan JE, Schnall MD, Schultz DJ, Solin LJ. Role of magnetic resonance imaging and magnetic resonance imaging-guided surgery in the evaluation of patients with early-stage breast cancer for breast conservation treatment. *Am J Clin Oncol* 1999; 22: 414–418.
10. Lee C. American Roentgen Ray Society; 2008.
11. DeMartini W, Lehman C. A review of current evidence-based clinical applications for breast magnetic resonance imaging. *Top Magn Reson Imaging* 2008; 19: 143–150.
12. *ACR BI-RADS: Breast Imaging and Reporting Data System*. Reston, VA: American College of Radiology; 2003.
13. Liberman L, Mason G, Morris EA, Dershaw DD. Does size matter? Positive predictive value of MRI-detected breast lesions as a function of lesion size. *AJR Am J Roentgenol* 2006; 186: 426–430.
14. LaTrenta LR, Menell JH, Morris EA, et al. Breast lesions detected with MR imaging: utility and histopathologic importance of identification with US. *Radiology* 2003; 227: 856–861.
15. DeMartini WB, Eby PR, Peacock S, Lehman CD. Utility of targeted sonography for breast lesions that were suspicious on MRI. *AJR Am J Roentgenol* 2009; 192: 1128–1134.

SPECIAL FEATURE

Current Surgical Management of Ulcerative Colitis

Alexis Grucela, MD, and Randolph M. Steinhagen, MD

¹Division of Colon and Rectal Surgery, Department of Surgery, Mount Sinai School of Medicine, New York, NY

OUTLINE

ULCERATIVE COLITIS

SURGICAL TREATMENT

Indications

Operative Management

Complications of

Ileal Pouch–Anal Anastomosis

Functional Results of

Ileal Pouch–Anal Anastomosis

Pouch Salvage Surgery

CONCLUSION

ABSTRACT

Inflammatory bowel disease is divided into 2 major disease entities: Crohn's disease and ulcerative colitis. Ulcerative colitis is characterized by contiguous inflammation of the colorectal mucosa, always beginning in and involving the rectum and progressing for variable distances proximally within the colon. In ulcerative colitis, medical therapy, which is not curative, is directed at controlling symptoms and reducing the underlying inflammatory process. However, emergent or elective removal of the colon and rectum does cure the disease and also eliminates the possibility of developing a malignancy. Here we present the current surgical treatment of ulcerative colitis and issues in the management of ulcerative

colitis. We discuss indications for surgical treatment, elective and emergent operative management, early and late complications of surgery, and functional results. *Mt Sinai J Med* 76:606–612, 2009. © 2009 Mount Sinai School of Medicine

Key Words: inflammatory bowel disease, restorative proctocolectomy, ulcerative colitis.

Inflammatory bowel disease is divided into 2 major disease entities: Crohn's disease (CD) and ulcerative colitis (UC). The etiology of the diseases is unknown. Both are characterized by intermittent episodes of exacerbation of their symptoms, and they have common clinical presentations and overlapping histological features.¹ However, as a result of the different distribution of these diseases within the intestinal tract, and their distinct behavior and nature of their inflammatory processes, the medical and surgical management of CD and UC are quite dissimilar. Here we present the current surgical treatment of UC and issues in the management of UC.

ULCERATIVE COLITIS

UC is characterized by contiguous inflammation of the colorectal mucosa, always beginning in and involving the rectum and progressing for variable distances proximally within the colon. This is distinctively different from Crohn's colitis, which often causes separate and distinct segments of colonic inflammation and may not involve the rectum.¹ In more extensive UC, patients may have complete involvement of the colon, or pancolitis. The incidence in the United States is roughly 9/100,000.² The cause is unknown. It occurs at all ages, with peak onsets between the second and fourth decades of life. The incidence of UC is equal among men and women, is more common in Jewish people, and is also

Address Correspondence to:

Randolph M. Steinhagen

Division of Colon and Rectal
Surgery Department of Surgery
Mount Sinai School of Medicine
New York, NY

Email: randolph.steinhagen@
mssm.edu

more common in whites than in nonwhites. Patients typically experience intermittent exacerbations with symptoms of bloody diarrhea associated with urgency and tenesmus. Inflammation in UC is confined to the mucosal and submucosal layers of the colon. This is an important contrast to CD, which is characterized by transmural inflammation of the bowel.²

SURGICAL TREATMENT

Indications

In UC, medical therapy is directed at controlling symptoms and reducing the underlying inflammatory process; this is not curative. However, removal of the colon and rectum does cure the disease and also eliminates the possibility of developing a malignancy, for which the most important risk factors are the duration and extent of the disease.³

In UC, medical therapy is directed at controlling symptoms and reducing the underlying inflammatory process; this is not curative. However, removal of the colon and rectum does cure the disease and also eliminates the possibility of developing a malignancy, for which the most important risk factors are the duration and extent of the disease.

In addition, surgery eliminates the need for patients to take anti-inflammatory medications on a long-term basis. Approximately 25% of patients with UC ultimately require colectomy.³ The indications for surgery can be divided into emergency and elective situations. Although surgery is completely curative, it is associated with risks and complications. These may include immediate postoperative issues possibly requiring reoperation, strictures and pouch functional problems, reduced fecundity in females, impotence and/or retrograde ejaculation in males, and pouch leakage and pelvic sepsis, which are the most morbid complications. Therefore, the surgeon must discuss all the risks and benefits with the patient before planning elective surgery.

Emergent indications include fulminant colitis unresponsive to medical therapy, toxic megacolon, perforation, and hemorrhage (Table 1).⁴ Fulminant colitis is the sudden onset of severe and frequent

Table 1. Indications for Surgery in Ulcerative Colitis.

Emergent	Elective
Fulminant colitis	Failure of medical therapy
Toxic megacolon	Dysplasia
Perforation	Malignancy
Hemorrhage	Stricture
	Unacceptable side effects of medications

(>10/day) bloody bowel movements, abdominal pain, dehydration, and anemia.⁴ Criteria for fulminant colitis, initially described by Truelove and Witts in 1955,⁴ include these symptoms and at least 2 of the following: tachycardia, a body temperature higher than 38.6°C (101.5°F), leukocytosis (>10,500/mm³), and hypoalbuminemia.^{4,5} Toxic megacolon occurs when the colon dilates significantly in the presence of fulminant colitis. Toxic megacolon may be the initial presentation of UC or may represent a flare in a patient with longstanding disease. Either the entire colon or an isolated segment of the colon (usually the transverse or left colon) can be involved.^{4,5} A radiographic definition of toxic megacolon is dilatation of the transverse colon greater than 5.5 to 6 cm on a supine abdominal film.^{4,5} However, toxic megacolon is truly a clinical diagnosis. Hemorrhage severe enough to result in hemodynamic instability is an unusual complication of UC.⁴ Elective indications for colectomy include

Emergent indications for colectomy include fulminant colitis unresponsive to medical therapy, toxic megacolon, perforation, and hemorrhage. Elective indications include symptoms and disease refractory to medical therapy, intolerance to or complications as the result of medicines, mucosal dysplasia, a dysplasia-associated lesion or mass, a malignancy, or the presence of a stricture.

symptoms and disease refractory to medical therapy, intolerance to or complications as the result of medicines, mucosal dysplasia, a dysplasia-associated lesion or mass, a malignancy, or the presence of a stricture (Table 1).¹ Because UC involves only the mucosa and submucosal layers of the colonic wall,

benign strictures are uncommon. When a stricture develops in a patient with UC, there must be heightened suspicion that there is an underlying malignancy. The increasing risk for the development of a malignancy in chronic UC with years of disease has been previously described.⁶ Alternatively, a benign stricture may indicate that the disease is actually CD (in which benign strictures are common) and not UC at all. In addition, there may be severe complications secondary to adverse effects of steroids or other medical therapy or growth retardation in children. Restorative proctocolectomy with ileal pouch–anal anastomosis (IPAA) has become the standard of care for patients with UC who ultimately require surgery.^{1,5}

Operative Management

Emergent Operative Management.

The aim of the operation in the emergency setting is to treat the fulminant colitis and restore the patient to his previous state of health in order to be able to perform a future restorative procedure. Pouch surgery should generally be avoided in the acute setting.⁷ The primary procedure performed for patients requiring emergent surgery for UC is total abdominal colectomy with an end ileostomy, with a rectal stump left behind. The advantage of this procedure is that it is

The primary procedure performed for patients requiring emergent surgery for UC is total abdominal colectomy with an end ileostomy, with a rectal stump left behind.

rapid and relatively safe, deals with the emergency, and restores the patient to good health while not disturbing the pelvic planes, avoiding the risk of pelvic bleeding and injury to the pelvic nerves and bladder.⁵ A future restorative operation can then be performed after the patient has fully recovered. The disease in the rectum left behind is inconsequential, and the vast majority of patients are able to discontinue all colitis medications, especially steroids.⁵ A mucus fistula can be created with the remaining rectosigmoid, but this is usually not recommended. Some surgeons leave a rectal tube inside the rectum for decompression and drainage in the immediate postoperative period, and some secure a longer rectosigmoid in the subcutaneous tissue at the lower end of the wound. This technique makes it much easier to locate the rectum at reoperation, and more importantly, stump dehiscence will result in a wound infection rather than an intraperitoneal leak or abscess.³

Elective Operative Management.

Restorative Proctocolectomy with Ileal

Pouch–Anal Anastomosis. Restorative proctocolectomy with IPAA has become the standard of care for elective surgery for UC. It was first described by

Restorative proctocolectomy with IPAA has become the standard of care for elective surgery for UC.

Parks and Nichols in 1978.⁸ It can be performed with open or laparoscopic approaches and has continued to evolve with the emergence of new technologies. There are 4 main steps to the procedure. First, the abdominal colon is removed. Then, pelvic dissection is performed, and the rectum is removed; the pelvic nerves and the anal sphincter are spared. Then, an ileal reservoir, most commonly a J pouch, is constructed. Finally, the pouch is anastomosed to the anal canal.⁸ The pouch–anal anastomosis is performed either with a double-staple technique or with a mucosectomy and hand-sewn anastomosis. Pouch designs have included the S, J, W, and H pouches.³ The majority of surgeons now favor the J pouch because of the facility of its construction, its economical use of the length of the terminal ileum, and its reliable emptying.³ The J pouch is created by the use of the distal 40 cm of the terminal ileum and by the joining of the antimesenteric borders with a linear cutting stapler, which divides the adjacent ileal walls, thereby creating a large chamber.³ This procedure has been shown to have an overall morbidity rate of 19% to 27% and a very low mortality rate of 0.2% to 0.4%. The quality of life following the operation has been shown to be near that of the normal population.¹

Laparoscopic Approaches. As advances have been made in laparoscopic technology, laparoscopic restorative proctocolectomy with IPAA has become more popular. Proponents of the laparoscopic approach feel that the benefits in comparison with open surgery are decreased postoperative pain, decreased length of hospitalization, and improved cosmesis.⁹ According to the data reported currently, the short-term and long-term complications, functional results, and quality of life are similar between laparoscopic and open groups.¹⁰ The length of hospitalization and need for postoperative pain medications are slightly better with the laparoscopic approaches. The single current documented benefit of laparoscopic IPAA is improvement in cosmesis.⁹ However, as the technology evolves and improves, future prospective studies may better delineate laparoscopic benefits.

Table 2. *Advantages of the Ileal Pouch–Anal Anastomosis Technique.*

Hand-Sewn Mucosectomy	Double-Stapled
All mucosa at risk removed	Less technically challenging
Precise placement of anastomosis	Faster
No cuffitis	Less risk of sphincter damage

Staged Procedures. Restorative proctocolectomy with IPAA can be performed as a single-stage operation in selected patients. It is most commonly performed as 2 stages, with the construction of a loop ileostomy at the same time as the construction of the ileal pouch. The purpose of the ileostomy is to divert the fecal stream and allow the pouch-anal anastomosis to fully heal. A subsequent operation a few months later is performed to close the loop ileostomy and restore intestinal continuity. The goal is to reduce the risk of abdominal sepsis and leakage.⁵ The rate of anastomotic leaks has been reported to range from 5% to 15% of cases.³ Proponents of

The rate of anastomotic leaks has been reported to range from 5% to 15% of cases.

this approach believe that the diverting ileostomy decreases the rate of pelvic sepsis.⁵ Supporters of a 1-stage procedure conversely believe that a nondiverted IPAA can be performed without an increased risk of pelvic sepsis. In addition to avoiding an ileostomy, a 1-stage procedure renders a second hospitalization and operation unnecessary, lowers the total cost, results in a shorter hospital stay, and perhaps even decreases the incidence of small bowel obstruction.^{11,12} Complication rates for ileostomy closure have been reported to range from 10% to 30%.³

Factors that are taken into account when it is being decided whether or not to stage the procedure include the overall condition of the patient, steroid use, age, and comorbidities.⁵ It is considered ideal when the anastomosis is intact on an intraoperative leak test under no tension, and the procedure is uncomplicated by factors such as high intraoperative blood loss.⁵ In a randomized control trial reported by Grobler *et al.*,¹³ there were 2 anastomotic leaks, 1 in the 2-stage group and 1 in the single-stage group. Some authors believe that there are comparable complication rates; however, some studies have shown that the severity of complications in the single-stage surgery is more severe.^{1,13}

Anal Anastomosis. The J pouch–anal anastomosis is performed either with a double-staple technique or with a mucosectomy with a hand-sewn anastomosis (Table 2). This still remains somewhat controversial. The goal of the mucosectomy is to remove all the disease or mucosa at risk by the resection of all the mucosa above the dentate line.^{14,15} This allows precise placement of the anastomosis at the dentate line.⁵ The disadvantages are that it is a more complex and technically challenging procedure and that it may predispose patients to higher rates of sphincter damage and incontinence.³ Additionally, there still may be a risk of residual colitic tissue as there have been reports of adenocarcinoma occurring in the anastomosis after a mucosectomy.¹⁵

The double-stapled method is technically less demanding, and those who advocate it at the level of the pelvic floor feel that the 1.5 to 2 cm of rectal mucosa proximal to the dentate line improves functional results by enhancing sensation.¹⁴ Long-term surveillance of any retained mucosa is required as there is the potential for dysplasia and cancer to develop.^{5,14,16} Another problem seen in association with a stapled anastomosis is symptomatic inflammation of the retained mucosa, a condition that has been called cuffitis. It is actually persistent UC in the short retained rectal segment.¹⁶ Mucosectomy is preferred for patients with multiple tumors or dysplastic regions or lesions close to the rectum. Pouch surgery is not advisable in the setting of low rectal cancer.¹⁶

Several studies have shown that pouch failure rates are the same with both types of anastomoses.⁵ In a prospective randomized trial from the Mayo Clinic, there was no significant difference in perioperative complications or functional results for patients who underwent mucosectomy with a hand-sewn IPAA versus patients with a double-stapled IPAA in which the anal transition zone was spared.¹⁷ However, there were higher resting sphincter pressures and less nocturnal incontinence in the stapled group, and this may indicate better long-term function.¹⁷

Total Proctocolectomy with Ileostomy.

Although restorative proctocolectomy with IPAA is the most common procedure performed in elective surgery for UC, total proctocolectomy with ileostomy is still performed for a select group of patients. It is the procedure of choice for patients who have impaired sphincter function, significant comorbidities, advanced age, or distal rectal cancer, or for those patients who elect not to have a restorative procedure.^{5,16} Complications can occur in up to 26% of patients and may include small bowel

obstruction, alterations in bladder function and sexual dysfunction, infertility, and, most frequently, stoma-related complications.¹⁶

Subtotal Colectomy with Ileorectal Anastomosis. Subtotal colectomy with ileorectal anastomosis may be an option in situations in which the rectum is minimally diseased. The ileum is anastomosed to the rectum, which must be distensible, must retain the capacity to act as a reservoir, and cannot have evidence of dysplasia or malignancy.⁵ Function is difficult to predict afterwards, and although some patients may experience multiple frequent stools and persistent rectal inflammation, some patients do quite well.¹⁸ Overall, approximately 25% of patients who undergo subtotal colectomy with ileorectal anastomosis will eventually require removal of the rectum because of persistent symptoms or the development of cancer.⁵ The advantage, however, is that pelvic dissection is avoided, and this eliminates the risk of impotence in males and reduced fecundity in females. As long as there is minimal rectal inflammation, the functional result of an ileorectal anastomosis is better than that of an ileal J pouch.

Proctocolectomy with a Kock Pouch (Continent Ileostomy). Another surgical option is total proctocolectomy with the creation of a Kock continent ileostomy. This procedure, which was popular in the 1970s before IPAA, was associated with a high rate of pouch dysfunction, which required surgical correction in the majority of cases. The Kock pouch is a continent ileostomy with a nipple valve that must be intubated for evacuation. The distal ileum is intussuscepted into a pouch constructed of more proximal bowel to form a continent valve.¹⁶ Today, it remains an alternative for patients who are not candidates for IPAA because of poor sphincter function, who do not want an ileostomy, or for whom IPAA has failed.¹⁶ It is rarely performed currently because of the high reoperation rate and the clear superiority of IPAA. The major problems seen are related to slippage (detussusception) of the nipple valve resulting in both incontinence and difficulty intubating the pouch to empty it.

Complications of Ileal Pouch—Anal Anastomosis

Early Complications.

Restorative proctocolectomy with IPAA is a technically challenging operation and has significant morbidity associated with it.

Restorative proctocolectomy with IPAA is a technically challenging operation and has significant morbidity associated with it.

The most commonly associated short-term complications include small bowel obstruction, anastomotic stricture, bleeding, pouch leakage, and pelvic abscess (Table 3). Early perioperative small bowel obstruction has been shown to occur in up to 15% of postoperative patients after IPAA. Although most respond to conservative management, some will ultimately require reoperation.¹⁹ An anastomotic stricture is present in one-third of patients who undergo reoperation after IPAA. These strictures can often be dilated in the office or the operating room under anesthesia.¹ The rate of postoperative pouch leakage and subsequent pelvic sepsis has been found to range from 5% to 15%.¹⁹ Some are treatable without surgery, with only intravenous antibiotics and computed tomography-guided drainage used. However, most require operative intervention and fecal diversion if that has not already been performed.¹⁹

Late Complications.

Long-term complications include small bowel obstruction, pouch fistulas, pouch dysfunction, pouchitis, sexual dysfunction, and reduced fecundity in women (Table 3). Pouchitis is the most frequent long-term complication of the procedure.^{1,5} This is believed to be an immunological response to the alteration of bowel flora in the pouch and is characterized by diarrhea and worsening of pouch function. Some patients have abdominal pain, fever, and symptoms similar to those of colitis.³ It has been reported in 12% to 50% of patients postoperatively, and some patients (5%–19%) require chronic therapy.²⁰ Metronidazole and ciprofloxacin are first-line therapy for pouchitis.³ There are very small numbers of patients who require pouch excision because of intractable pouchitis.²⁰ Fistulas are difficult to treat and include pouch-vaginal and pouch-perianal fistulas. Although

Table 3. Complications After Ileal Pouch—Anal Anastomosis.

Short Term	Long Term
Small bowel obstruction	Small bowel obstruction
Stricture	Pouch fistula
Bleeding	Pouch dysfunction
Leak	Pouchitis
Pelvic abscess	Sexual dysfunction
	Reduced fecundity

early fistulas likely are the result of imperfect handling of the IPAA, a late fistula should raise concern for CD.¹

The issue of fecundity after IPAA was recently investigated in a Swedish study.²¹ Previous studies indicated that there was a significant reduction in postoperative fertility in females with UC after IPAA. The birth rate in women with UC was similar to that in age-matched women who had not undergone surgery. There was a significant decrease in births in women with UC after IPAA, and many more of those births were a result of in vitro fertilization: 29% versus the expected 1% in vitro fertilization birth rate in Sweden.²¹ A recent study from our own institution showed similar results.²² The cause of decreased fecundity after IPAA is most likely scarring in the pelvis, which makes it more difficult for the egg to reach the oviduct after ovulation. Anatomic changes in the pelvis may also contribute.

Functional Results of Ileal Pouch–Anal Anastomosis

Similar functional results have been obtained by many different surgeons at different institutions. Most patients report good to excellent pouch function.¹ The most commonly used markers of pouch function

Most patients report good to excellent pouch function.

are the number of diurnal and nocturnal bowel movements, the number of episodes of soiling, the number of episodes of incontinence, and the extent to which medications are required to control bowel activity.¹

In a large study at the Mayo Clinic, Farouk *et al.*²⁰ found that, in 1386 patients after IPAA for UC with a median follow-up of 8 years, pelvic sepsis was the primary cause of pouch failure, regardless of sex or age. Incontinence rates were significantly higher in older patients versus younger patients. However, this did not result in a greater risk of pouch failure in the older patients. Nearly 80% of patients had complete diurnal continence, whereas 19% had occasional episodes of incontinence, and 2% had frequent episodes.²⁰ Nearly 50% of patients were discharged from the hospital on some type of medication to slow their bowels or to provide dietary bulk. In patients who had an ileal pouch for more than 10 years, stool frequency and continence remained remarkably stable over time.²⁰

The risk of long-term pouch loss has been reported to range from 1% to 20% in different studies.⁴ The overall rate of pouch loss is thought

to be approximately 10%. When there is a problem, management options, depending on the nature of the problem, include diverting ileostomy, pouch excision and end ileostomy, and pouch revision.¹

Pouch Salvage Surgery

Pouch failure rates in IPAA for UC have been found to range from 3% to 15%.²³ A permanent diverting ileostomy with or without excision of the pouch is no longer the only alternative for patients with pouch failure. Salvage procedures for pouch failure now include transanal pouch mobilization and advancement as well as abdominoperineal pouch reconstruction or revision, with or without neoileal pouch–anal anastomosis. These operations have been reported in the literature as having the best outcomes and salvage rates for failed pouches.²³ Success rates have been reported to range from 48% to 93%.²⁴ Perineal procedures are usually attempted first, as they are less invasive and have a slightly higher success rate. However, the procedure choice is most dependent on the etiology and anatomy.²⁴ Both functional and quality of life outcomes have been found to be good after pouch salvage surgery.²³

An additional alternative for patients with pouch failure is continent ileostomy, or Kock pouch creation. This is a possibility in highly motivated patients who wish to avoid an external ileostomy appliance after pouch failure and who are not candidates for a redo pouch procedure. This is applicable only to a highly select group of patients at specialized centers. Sometimes, the original pelvic pouch can be used to create the Kock pouch. Lei *et al.*²⁵ performed a review of patients who underwent continent ileostomy creation after pouch failure. In the patient populations studied, the morbidity associated with this procedure was high. The results initially seemed poor as the long-term dysfunction rate was 50%, the complication rate was 60.9%, and the revision rate was 45.3%. However, the pouch retention rate was very good (95%), as was the postoperative quality of life.²⁵

CONCLUSION

IPAA has revolutionized the standard of treatment for UC. The operation is safe and resilient with an acceptable morbidity rate (19%–27%) and a low mortality rate (0.2%–0.4%).⁵ Surgery for UC is curative,

IPAA has revolutionized the standard of treatment for UC.

whereas that for CD is still only palliative, with a high likelihood of recurrence. Although the 2 disease entities present similarly, their behavior and natural history result in very different surgical goals and management.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Cima R, Pemberton JH. Medical and surgical management of chronic ulcerative colitis. *Arch Surg* 2005; 140: 300–310.
2. Loftus CG, Loftus EV Jr, Harmsen WS, et al. Update on the incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota, 1940–2000. *Inflamm Bowel Dis* 2007; 13: 254–261.
3. Bach SP, Mortensen NJ. Ileal pouch surgery for ulcerative colitis. *World J Gastroenterol* 2007; 13: 3288–3300.
4. Truelove SC, Witts LF. Cortisone in ulcerative colitis: final report on a therapeutic trial. *BMJ* 1955; 2: 1041–1048.
5. Cohen JL, Strong SA, Hyman NH, et al. Practice parameters for the surgical treatment of ulcerative colitis. *Dis Colon Rectum* 2005; 48: 1997–2009.
6. Gorfine SR, Bauer JJ, Harris MT, et al. Dysplasia complicating chronic ulcerative colitis: is immediate colectomy warranted? *Dis Colon Rectum* 2000; 43: 1575–1581.
7. Becker JM. Surgical therapy for ulcerative colitis and Crohn's disease. *Gastroenterol Clin North Am* 1999; 28: 371–390.
8. Parks AG, Nichols RJ. Proctocolectomy without ileostomy for ulcerative colitis. *BMJ* 1978; 2: 85–88.
9. Dunker MS, Bemelman WA, Slors JFM, et al. Functional outcome, quality of life, body image, and cosmesis in patients after laparoscopic-assisted and conventional restorative proctocolectomy: a comparative study. *Dis Colon Rectum* 2001; 44: 1800–1807.
10. Sylla P, Chessin DB, Gorfine SR, et al. Evaluation of one-stage laparoscopic-assisted restorative proctocolectomy at a specialty center: comparison with the open approach. *Dis Colon Rectum* 2009; 52: 394–399.
11. Sugerman HJ, Sugerman EL, Meador JG, et al. Ileal pouch anal anastomosis without ileal diversion. *Ann Surg* 2000; 232: 530–541.
12. Heuschen UA, Hinz U, Allemeyer EH, et al. One- or two-stage procedure for restorative proctocolectomy: rationale for a surgical strategy in ulcerative colitis. *Ann Surg* 2001; 234: 788–794.
13. Grobler SP, Hosie KB, Keighley MRB. Randomized trial of loop ileostomy in restorative proctocolectomy. *Br J Surg* 1992; 79: 903–906.
14. Remzi FH, Fazio VW, Delaney CP, et al. Dysplasia of the anal transitional zone after ileal pouch-anal anastomosis: results of prospective evaluation after a minimum of ten years. *Dis Colon Rectum* 2003; 46: 6–13.
15. Rodriguez-Sanjuan JC, Polavieja MG, Naranjo A, et al. Adenocarcinoma in an ileal pouch for ulcerative colitis. *Dis Colon Rectum* 1995; 38: 779–780.
16. Metcalf AM. Elective and emergent operative management of ulcerative colitis. *Surg Clin North Am* 2007; 87: 633–641.
17. Reilly WT, Pemberton JH, Wolff BG, et al. Randomized prospective trial comparing ileal pouch-anal anastomosis performed by excising the anal mucosa to ileal pouch-anal anastomosis performed by preserving the anal mucosa. *Ann Surg* 1997; 225: 666–676.
18. Bellolio RF, Zúñiga AJM, Wagner HP, et al. Ileorectal anastomosis in the surgical treatment of ulcerative colitis: long-term results. *Rev Med Chil* 2008; 136: 1121–1126.
19. Meagher AP, Farouk R, Dozois RR, et al. J ileal pouch-anal anastomosis for chronic ulcerative colitis: complications and long-term outcome in 1310 patients. *Br J Surg* 1998; 85: 800–803.
20. Farouk R, Pemberton JH, Wolff BG, et al. Functional outcomes after ileal pouch-anal anastomosis for chronic ulcerative colitis. *Ann Surg* 2000; 231: 919–926.
21. Ording Olsen K, Juul S, Berndtsson I, et al. Ulcerative colitis: female fecundity before diagnosis, during disease, and after surgery compared with a population sample. *Gastroenterology* 2002; 122: 15–19.
22. Chessin DB, Gorfine SG, Bub DS, et al. The Impact of Restorative Proctocolectomy on Fecundity: Predictors of a Successful Pregnancy. Presented at Annual Meeting of the American Society of Colon and Rectal Surgeons; June 2008; Boston, MA.
23. Remzi FH, Fazio VW, Kirat HT, et al. Repeat pouch surgery by the abdominal approach safely salvages failed ileal pelvic pouch. *Dis Colon Rectum* 2009; 52: 198–204.
24. Shawki S, Belizon A, Person B, et al. What are the outcomes of reoperative restorative proctocolectomy and ileal pouch-anal anastomosis surgery? *Dis Colon Rectum* 2009; 52: 884–890.
25. Lei L, Fazio VW, Remzi FH, et al. Outcomes for patients undergoing continent ileostomy after a failed ileal pouch-anal anastomosis. *Dis Colon Rectum* 2009; 52: 1409–1414; discussion 4414–4416.

SPECIAL FEATURE

Genomics of Cardiovascular Disease

Leslie Cole Manace,¹ Tapan N. Godiwala,² and Mark W. Babyatsky²¹Department of Genetics, Kaiser Oakland Medical Center, Oakland, CA²Department of Medicine, Mount Sinai School of Medicine, New York, NY

OUTLINE

INTRODUCTION

CONTRIBUTION OF INHERITED CARDIOVASCULAR
SYNDROMES

CURRENT GENOMIC RESEARCH

METHODOLOGY AND
KEY ASSOCIATIONSAPPLICATIONS OF GENOME-WIDE ASSOCIATION
STUDIES

Tailored Drug Therapy

Diagnostic and

Prognostic Genomic Profiles

SUMMARY

ABSTRACT

As the leading cause of death worldwide and a major cause of disability, cardiovascular disease remains a central focus of basic research, pharmacological treatment, surgical interventions, and long-term care. Inherited, monogenic syndromes have provided insight into pathophysiological mechanisms across the range of cardiovascular diseases. With the advent of post-Human Genome Project resources and technology, there has been a flood of research aimed at genome-wide predisposition markers, pharmacogenetics, and genomic signatures in complex cardiovascular disorders. Genomic research has both further elucidated the impact of genes previously identified in cardiovascular disease development and progression and discovered genomic regions as yet unknown to be associated with cardiovascular outcomes. The promise of personalized medicine

lies in combining this genetic information with other biomarkers to tailor preventive and therapeutic strategies to individual patients for effective management, fewer adverse events, and preventive care. *Mt Sinai J Med* 76:613–623, 2009. © 2009 Mount Sinai School of Medicine

Key Words: cardiology, genomics, pharmacogenetics.

A 74-year-old woman presented to an outside hospital's emergency department with complaints of chest pain. She reported 6 hours of substernal chest pressure radiating to the neck and bilateral upper extremities accompanied by severe shortness of breath, diaphoresis, and palpitations consistent with her previous presentation of acute myocardial infarction (MI). An electrocardiogram (ECG) revealed 3-mm ST elevations in V2 and V3 and T-wave inversions in the lateral leads. The patient was started on a heparin drip, was given sublingual nitroglycerin, aspirin (325 mg), clopidogrel (600 mg), and furosemide (40 mg), and was transferred to our hospital for further evaluation and treatment.

Her past medical history included systolic heart failure (ejection fraction of 30%), childhood rheumatic fever, diabetes, and hypertension. Fourteen years ago, the patient had radiofrequency ablation for intractable atrial arrhythmias. Two years prior to admission, she was found to have a positive stress test during the workup for dyspnea on exertion. She underwent 2 staged percutaneous interventions for 3-vessel coronary artery disease (CAD); a total of 3 drug-eluting stents were placed in the second diagonal branch of the left anterior descending (LAD) coronary artery, in the mid-LAD coronary artery, and in the first obtuse marginal branch of the left circumflex arteries. Four months later, the patient presented to a hospital in Cape Cod with thrombosis of 1 of her LAD stents, requiring urgent balloon angioplasty. Two months prior to admission, an automatic implanted cardiac defibrillator was placed for primary prevention because of her poor ejection fraction.

Address Correspondence to:

Leslie Cole Manace

Department of Genetics
Kaiser Oakland Medical Center
Oakland, CA

Email: leslie.c.manace@kp.org

The patient was of Hispanic background. Her family history was significant for heart disease: her mother and brother both died of MI at the ages of 70 and 62 years, respectively, and her father died of colon cancer. The patient reported a distant 12-pack-year smoking history. Illicit drug or alcohol abuse was denied.

Medications included sublingual nitroglycerin as needed, clopidogrel (75 mg daily), aspirin (81 mg daily), furosemide, warfarin, metoprolol, isosorbide mononitrate, pantoprazole, glipizide, and sitagliptin. Previous statin therapy was discontinued because of myalgias. The patient stated that she was fully compliant with her antiplatelet regimen.

On physical examination, the patient was tachypneic and slightly distressed though alert and oriented. The patient's temperature was 35.8°Celsius, her heart rate was 101 beats per minute, her blood pressure was 115/68 mm Hg, her respiratory rate was 30 breaths per minute, and her oxygen saturation was 95% on 2 L of nasal oxygen. Her physical examination was remarkable only for bilateral rales on pulmonary examination.

The first set of cardiac troponins was negative at the outside hospital. Labs checked at this institution revealed a troponin level of 10.9 ug/L, a creatine kinase MB level of 10.6 U/L, a white blood cell count of 14.4×10^3 with neutrophil predominance, an international normalized ratio of 1.7, and a partial thromboplastin time of 149 seconds (on heparin). A chest X-ray revealed mild venous congestion. Repeat ECGs lacked the previously seen ST elevations but did show continued T-wave inversions in the lateral leads.

Approximately 4 hours after the initiation of therapy at the outside hospital, the patient arrived at the coronary care unit of this institution, where an abciximab drip was started. Cardiac catheterization revealed thrombosis of the diagonal and mid-LAD stents. Balloon angiography was successfully performed, and an intra-aortic balloon pump was placed.

On hospital day 2, the heparin drip was stopped, and the intra-aortic balloon pump was removed. On hospital day 3, the patient was transferred to the telemetry unit and subsequently discharged home. Genotyping for vitamin K epoxide reductase complex subunit I (*VKORC1*), cytochrome P450 2C9 (*CYP2C9*), and cytochrome P450 2C19 (*CYP2C19*) was sent. The patient was heterozygous for the hepatic enzyme gene *CYP2C19* (i.e., she possessed different versions of the gene on the maternal and paternal chromosomes), with 1 wild-type allele (*CYP2C19*1*) and a variant (*CYP2C19*2*) representing the 681G>A polymorphism, which leads to reduced enzyme activity. Both *CYP2C9* alleles were wild-type. The patient was also heterozygous for the gene *VKORC1*, which

is involved in recycling vitamin K and affects warfarin activity, with 1 wild-type *VKORC1* allele and a variant (*VKORC1*2*) representing a common promoter (ie, a regulatory region directly before a gene), the 1639G>A polymorphism. On the basis of the genotyping data, the patient was predicted to be an intermediate metabolizer for drugs processed by *CYP2C19* (including clopidogrel) and for drugs processed by *CYP2C9* (including warfarin) and to be warfarin-sensitive and thus require reduced dosing.

INTRODUCTION

As the leading cause of death worldwide, cardiovascular disease remains a central focus of basic research, pharmacological treatment, surgery, and chronic care. Monogenic cardiac diseases have provided insight into pathophysiological mechanisms and the genetic underpinnings of a growing number of complex cardiovascular disorders that are caused by interactions between multiple genes and environmental factors. Since the sequencing of the human genome, resources and technology have been directed toward genome-wide predisposition markers, pharmacogenetics, and genomic signatures. The promise of personalized medicine lies in combining this genetic information with other biomarkers to tailor preventive and therapeutic strategies to individual patients for improved efficacy and fewer adverse effects. We will discuss here some key concepts and specific discoveries of genomic research in cardiology, the unique suitability of cardiovascular disease for genomic personalized medicine, emerging treatment strategies in cardiology incorporating genetic information, and how genomics has already enhanced our basic understanding of cardiovascular disease and its clinical prevention and management (Table 1).

Genomics has already enhanced our basic understanding of cardiovascular disease and its clinical prevention and management.

CONTRIBUTION OF INHERITED CARDIOVASCULAR SYNDROMES

After elucidating certain pathways in cardiovascular biology and pathology with monogenic diseases (a single gene mutation leading to disease with

Table 1. Genetic Predisposition and Pharmacogenetic Associations in Cardiovascular Disease.

Genomic Locus	Disease or Pharmacological Association
Potassium channel <i>KCNQ1</i> and <i>KCNH2</i> genes	Mutations cause long QT1 syndrome and long QT2 syndrome, respectively, with different eliciting factors and treatment recommendations. Both carry a risk of sudden cardiac death.
Sodium channel <i>SCN5A</i> gene	Mutations lead to long QT3 syndrome, Brugada syndrome, or both through defects in cardiac sodium ion channels. There is a predisposition to atrial fibrillation.
9p21 region	There is an association with CAD and MI as well as intracranial and aortic aneurysms.
4q25 region	There is an association with atrial fibrillation.
Factor V Leiden (<i>F5</i>) and prothrombin (<i>F2</i>) genes	Polymorphisms R506Q and 20210G>A, respectively, in these coagulation factors result in an inherited hypercoagulable state. Testing for factor V Leiden is indicated for venous thrombosis in any individual younger than 50 years or in unusual sites.
<i>CYP2C19</i> gene	Loss-of-function alleles result in diminished conversion of clopidogrel to its active metabolite, increasing the risk for major CV events and coronary stent thrombosis.
<i>CYP2C9</i> and <i>VKORC1</i> genes	<i>CYP2C9</i> gene polymorphisms affect the metabolism of warfarin in the liver. <i>VKORC1</i> gene polymorphisms lead to differential rates of vitamin K recycling. Both have profound effects on warfarin dosing.

Abbreviations: CAD, coronary artery disease; CV, cardiovascular; CYP2C19, cytochrome P450 2C19; CYP2C9, cytochrome P450 2C9; F2, coagulation factor II; F5, coagulation factor V; KCNH2, potassium voltage-gated channel subfamily H (eag-related) member 2; KCNQ1, potassium voltage-gated channel KQT-like subfamily member 1; MI, myocardial infarction; SCN5A, sodium channel voltage-gated type V alpha subunit; VKORC1, vitamin K epoxide reductase complex subunit I.

a clear Mendelian inheritance pattern), we identified important molecular players and interconnections. From an inborn error of amino acid metabolism, homocysteinuria, the detrimental effect of high circulating homocysteine levels in promoting thrombosis, was discerned. Subsequently, genes with common population variants, such as *MTHFR* which encodes 5,10-methylenetetrahydrofolate reductase involved in homocysteine metabolism, as well as vitamins including folic acid and B12, which act as cofactors in the homocysteine pathway, have been examined to determine whether they influence the risk of cardiovascular complications.^{1,2}

Families with autosomal dominant hypercholesterolemia have demonstrated the impact of high low-density lipoprotein cholesterol levels on CAD.

Families with autosomal dominant hypercholesterolemia have demonstrated the impact of high low-density lipoprotein cholesterol levels on CAD.

Conversely, mutations in the glycoprotein gene *PCSK* result in low levels of low-density lipoprotein and a decreased prevalence of ischemic heart disease, even in the presence of other conventional risk

factors for atherosclerosis.³ New insight disputing the protective role of the high-density lipoprotein fraction of cholesterol in ischemic heart disease was also aided by monogenic conditions. In Tangier's disease, which is characterized by mutations in a cholesterol efflux pump gene *ABCA1* that correlate with low or absent levels of high-density lipoprotein and massive tissue deposition of cholesterol esters, there is no increased risk of ischemic heart disease if the patient's low-density lipoprotein fraction and triglyceride levels are normal; this suggests that low high-density lipoprotein levels alone may not be a risk factor for CAD.⁴ Supporting this finding, mutations in the cholesteryl ester transfer protein gene yield higher high-density lipoprotein levels but are not protective against ischemic heart disease.⁵ Building on the identification of these genes and the pathological significance of their protein products, more recent studies have assessed gene variants across the genome (from regions reaching over all 23 pairs of chromosomes) for correlations with ischemic heart disease risk. This year, a large study demonstrated that 9 genomic sites (or loci) are significantly associated with early-onset MI.⁶ Further work should continue to unravel the pathogenesis of CAD in terms of both individual gene contributions and broader molecular pathways.

Monogenic diseases also demonstrate gene-gene and gene-environment interactions contributing to cardiovascular pathology. The familial hypertrophic

Monogenic diseases also demonstrate gene-gene and gene-environment interactions contributing to cardiovascular pathology.

cardiomyopathies, caused by mutations in genes encoding components of the sarcomere, are characterized by left ventricular hypertrophy in the absence of predisposing cardiac conditions and exemplify allelic and locus heterogeneity (the similar clinical presentations are due to different mutations in the same gene or in distinct genes, respectively). It is important to identify Mendelian hypertrophic cardiomyopathies for the screening of family members, prevention, and appropriate therapy. Specific gene panels designed to assess patients for causative hypertrophic cardiomyopathy mutations and arrive at the proper molecular diagnosis are discussed later.

Long QT syndrome is one of the most common genetic arrhythmia conditions. Hundreds of

Long QT syndrome is one of the most common genetic arrhythmia conditions.

mutations have been described to date in 10 genes, with the long QT syndrome 1 (LQT1) to LQT3 subtypes accounting for 90% of cases. Long QT syndrome typically presents as unexplained syncope, palpitations, or sudden cardiac death in young adults with structurally normal hearts. Triggers for symptoms or cardiac events include a long list of prescription and over-the-counter medications, loud noises, exercise, and intense emotions. Transmission is usually autosomal dominant [only 1 mutation on an autosome (chromosomes 1–22 and not sex chromosome X or Y) causes the syndrome] or recessive (mutations in both maternal and paternal alleles are necessary to lead to disease), although not all cases are strictly Mendelian. Furthermore, penetrance of the phenotype is variable (not all carriers of the mutated allele express clinical evidence of the disease). The most common subtype, LQT1, results in a loss-of-function mutation in the gene potassium voltage-gated channel *KCNQ1* which encodes a potassium channel. The LQT2 subtype is the result of a loss-of-function mutation (reduced or complete loss of the normal function of the encoded

protein) in the gene potassium voltage-gated channel subfamily H (eag-related) member 2, which encodes another potassium channel. Finally, mutations in a cardiac sodium channel, sodium channel voltage-gated type V alpha subunit (*SCN5A*), result in the LQT3 phenotype. The sensitivity of genetic testing is approximately 50% to 75% in those patients tested with a high clinical suspicion of long QT syndrome. There are multiple applications for genetic testing in this syndrome: treatment can be tailored according to the specific mutation identified with appropriate pharmaceuticals or avoidance of the recognized trigger, and the testing of family members of an affected proband (the patient first identified with the syndrome and/or a causative mutation) can confirm or exclude the diagnosis and guide surveillance and prevention.⁷

Brugada syndrome illustrates the human genetic principles of variable expressivity and penetrance. As the second leading cause of death of Asian males less than 40 years old, the syndrome is not rare and is clinically relevant in routine cardiology. Characteristic ST-segment elevations in leads V1 to V3 may be found on ECGs, although these ECG findings are often absent or appear only if provoked by tachycardia or specific medications. Similar to long QT syndrome, Brugada syndrome increases the risk of ventricular arrhythmias and sudden death, although the disease manifests with a spectrum of severity (variable expressivity). The first gene mutation associated with Brugada syndrome was the sodium channel gene *SCN5A*. A specific variant of *SCN5A* is present in 13% of African Americans, although it may not be clinically detected in the same proportion (variable penetrance).^{8,9} Because sudden cardiac arrest is a major cause of mortality in Western populations on the whole, unmasking Brugada syndrome (which may require the use of provoking drugs during an ECG study) or associated genetic variants highlights the importance of accurate molecular diagnosis in the clinical management of patients.¹⁰ Recently, mutations in *SCN5A* have been associated with either Brugada syndrome or LQT3, depending on specific clinical contexts such as the heart rate.¹¹ *SCN5A* mutations have also been linked to atrial fibrillation. Other familial mutations in cardiac potassium channel genes result in a subset of patients with atrial fibrillation, and this further implicates ion channels in the pathogenesis of this common arrhythmia.

CURRENT GENOMIC RESEARCH METHODOLOGY AND KEY ASSOCIATIONS

Going beyond what insight Mendelian diseases can provide in cardiovascular pathophysiology, genome-wide association studies (GWASs) grew logarithmically over the last decade, taking advantage of the genomic data and technology available in the wake of the Human Genome Project. GWASs compare large populations with common, complex diseases in case-control studies discovering genetic markers that align with a disease state to a statistically significant degree. Although traditional genetic research examined individual genes with profound phenotypic effects through the use of families with multiple affected members or with a hypothesis-driven model of biological plausibility for the gene product, genome-wide research evaluates genetic markers across the genome in a hypothesis-independent manner or within specific candidate regions to discover new loci associated with common, complex (multifactorial) traits.¹² Genomic

Genome-wide research evaluates genetic markers across the genome in a hypothesis-independent manner or within specific candidate regions to discover new loci associated with common, complex (multifactorial) traits.

research takes advantage of high-throughput, highly multiplexed sequencing technologies developed alongside the Human Genome Project, as well as the torrent of DNA sequence data produced by that and subsequent projects, to identify genetic loci that are statistically linked to the outcome under study. We have found reproducible associations across the genome with cardiovascular disease development, and the challenge in the coming years will be to discern the specific roles that these markers play in pathogenesis: Do the gene products participate in biological pathways of the disease process or in the regulation of pathway intermediates? Is the marker merely close to the gene of actual interest? Are there stochastic gene-gene interaction effects of the marker on other close or distant genes? It must be emphasized that genetic association alone does not indicate biological causation; only with laborious protein studies and animal model work can the

We have found reproducible associations across the genome with cardiovascular disease development, and the challenge in the coming years will be to discern the specific roles that these markers play in pathogenesis.

role of a gene or DNA region identified by GWASs be explored. Additionally, the genetic marker must be considered in the systems biology context of environmental, genetic, and metabolic factors that may modify its influence on cardiovascular disease progression.

Because GWASs are geared to discover common variants, most have small phenotypic effects (or penetrance) and thus confer very modest odds ratios, the risk estimates of choice in case-control studies. Constructing a properly powered GWAS to detect variants with a small effect requires large and preferentially heterogeneous populations, molecular genetic processing, and massive data analysis, and this often necessitates cooperation between centers. One of the benefits of genomic research is this collaborative nature of the enterprise, one that mirrors the international cooperation that produced the draft human genome sequence. However, significant work following GWASs to confirm the strength of genomic marker association across multiple ethnicities and subcategorizations of cardiovascular disease is needed. In order for GWAS findings to be sufficiently interrogated and potentially incorporated into medical care, consortia-run prospective clinical studies will be critical to establish the clinical validity and utility (whether the DNA sequence result bears merit in identifying a particular clinical state and whether the test is cost-effective and beneficial, respectively) of these markers.

A roster of reproducible genetic associations in cardiovascular disease obtained via GWASs is currently the subject of much research attention. Until we can inexpensively sequence the full genomes of study participants (this may arrive in a handful of years, given the exponential progress of sequencing technology), we will use hundreds of thousands to millions of genomic loci to assess an association with disease development or outcome. Many GWASs use variations in single base pairs that are present in at least 1% of the population and distributed across the genome, which are known as single nucleotide polymorphisms (SNPs), to obtain genetic linkage. Each person has tens of thousands of

stable, usually inherited variants in both coding and noncoding portions of the genome that are not known to cause drastic changes in the function or amino acid codon of the area (silent mutations) but are useful because, as variants between people and populations, the SNPs move in blocks with genes they are in or nearby and identify particular versions of a whole genomic region or haplotype. For example, an SNP that is a cytosine in one person and a guanine in another may be in or next to a gene of interest for a certain disease and may be found to travel more in people with atrial fibrillation (with the cytosine variant) than in those without it (with the guanine variant) (Fig. 1). Until sequencing technology becomes efficient and inexpensive to the point that we can rapidly and fully sequence subjects' genomes, the HapMap takes advantage of the patterns of SNPs in blocks (haplotypes) to form a template for genetic discovery. GWASs thus produce probabilistic results that populations with a certain variant have a higher risk of developing a disease or going on to a certain outcome. It is important to distinguish that these results (1) are not deterministic for individuals; (2) are specific to the populations studied (largely limited to European backgrounds thus far); and (3) do not provide information about the biological role or other modifying factors for the gene or genomic locus.¹³

One genome locus that has been repeatedly associated with CAD and an increased risk of MI is 9p21 (chromosome 9, short p arm, band 2, subdivision 1), which refers to dark heterochromatin-unexpressed regions revealed by chromosome staining. In 2007, 4 groups described a strong association between CAD/MI and variants in the 9p21 region with a per allele copy odds ratio of approximately 1.25.^{14–17} The correlation has been replicated and expanded to include abdominal aortic aneurysm and intracranial aneurysm, providing an intriguing commonality to disorders of blood vessel walls.^{6,18} The most predictive variant, an SNP designated rs1333049 that encodes the cytosine allele lying in the 9p21.3 locus, has a yet unknown function, although its distance from 2 tumor suppressor genes (inhibiting cell growth) is small, and this may be related to its role. It has been more recently shown that the variant appears to be associated with disease prevalence and not severity^{19,20}; however, there is evidence collected from a prospective study that carrying the 9p21 cytosine allele leads to worse atherosclerotic burden.²¹

Not yet having outlined the 9p21 locus's biological relevance, frequency in a diverse set of populations, or susceptibility to other genetic and environmental modifiers, we have much research

to be do before we can test individual patients and know the precise impact of this variant on their cardiovascular risk calculation. It is plausible that it may provide important preventive information in the case of abdominal aortic or intracranial aneurysms that otherwise may not be suspected until a catastrophic, late clinical presentation. Whether the cost of increased medical examinations and patient apprehension will be outbalanced by the benefit of such predictive tests remains to be determined. A recent GWAS examining 9 loci previously associated with a risk for CAD showed that the aggregate of gene tests was associated with acute MI, although the loci were not shown to be individually associated.⁶

Genetic determinants of atrial fibrillation—the most common sustained cardiac arrhythmia with a prevalence as high as 1 in 10 individuals over 80 years of age—have been increasingly uncovered by GWASs. Large groups such as the deCODE genetics team and an international consortium including the Framingham Heart Study identified variants in locus 4q25 across European and Asian populations with atrial fibrillation with odds ratios between 1.36 and 1.90.^{22,23} The variants are adjacent to the gene *PITX2* which is known to have a role in the development of right-left asymmetry of the heart.²² Other studies have begun with a candidate gene known to function in heart development or rhythm and have examined SNPs within that gene for association with disease outcome. In this manner, 2 variants in the promoter region (upstream or before the coding region of a gene, where molecules may bind and affect gene transcription) of connexin 40, which is known to be important for electrical coupling between atrial myocytes, were found to be significantly associated with atrial fibrillation, with an odds ratio up to 2.53.²⁴

Given the impact of elevated blood pressure on cardiovascular disease and mortality, the discovery of genomic associations with hypertension has been possible with the meta-analysis of ever larger study populations and the expansion of SNP coverage across the genome. The largest trial to date identified 8 variants significantly associated with systolic or diastolic blood pressure and with hypertension near known genes.²⁵ One SNP was located in an intron of *MTHFR*, a gene participating in homocysteine metabolism, as discussed previously, as well as blood pressure regulation. It is important to evaluate loci and genes near the associated SNP as any SNP or combination of SNPs within the haplotype block may be truly involved in disease pathogenesis or progression. The associated SNP may be in the intron of *MTHFR*, but that does not imply a pathological correlation with the *MTHFR* product; adjacent genes

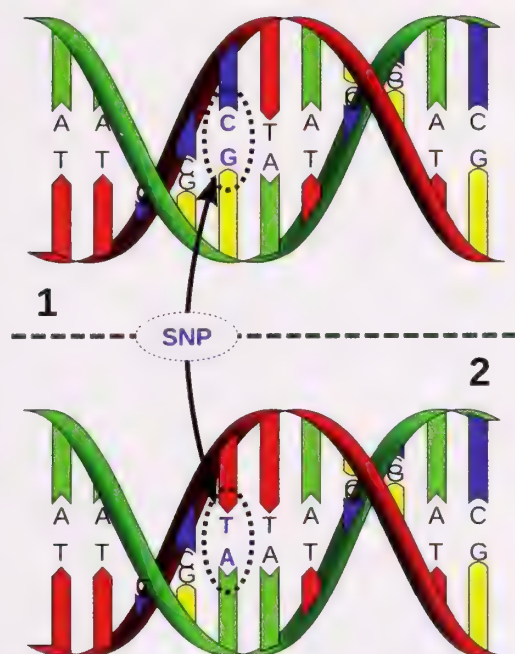


Fig 1. Single nucleotide polymorphism (SNP). Reprinted with permission from David Hall (licensed under Creative Commons Attribution 2.5).

(in this case, *CLCN6*, *NPPA*, *NPPB*, and *AGTRAP*) are also plausible candidates for actual biological importance.

Venous thromboembolism is another good candidate for the GWAS approach as a common, complex disease with a genetic component not well delineated beyond a few known loci. Relatively common variants in the genes for factor V (factor V Leiden, in which the R506Q SNP changes the arginine amino acid encoded at protein position 506 to a glutamine) and prothrombin (in which the G20210A SNP changes the nucleotide guanine to adenine at base pair position 20210) have been known for some time to increase the risk of initial and recurrent venous thromboembolism, with clear pathological mechanisms in clot formation.²⁶

APPLICATIONS OF GENOME-WIDE ASSOCIATION STUDIES

Tailored Drug Therapy

Pharmacogenetics refers to the impact of genetic variation on the metabolism, efficacy, and toxicity of drugs between individuals. There are families of metabolic enzymes known to participate in the breakdown of most drugs in the liver, notably including the cytochrome P450 (CYP450) system, that

have become the focus of many pharmacogenetic studies. Additionally, gene products specific to each drug's mechanism of action have also come to the forefront. The frontier in pharmacogenetic research is now the biological relevance of the markers discovered via GWASs and the clinical utility of including genetic testing in drug selection and dosing.

Pharmacological treatment is a mainstay of cardiovascular disease management; indeed, the highest grossing drugs internationally target heart disease and have provided large decreases in morbidity and mortality. Although for many components of cardiovascular disease, from dyslipidemia to hypertension to thrombophilia, there are effective medication options, the variability in interindividual response and especially adverse events is of continued concern. By incorporating pharmacogenetic results characterizing genetic variants involved in drug breakdown and reaction, we may more effectively arrive at medication selection and dosage and limit adverse reactions.

As illustrated in the case report, the treatment of acute coronary syndromes has been revolutionized with the advent of coronary artery stenting. Although these stents have been instrumental in care for patients with anginal symptoms, stenosed vessels, and impending MI, restenosis of stents has become a major issue both acutely after stent placement and in the long-term management of this large group of patients. Genetic determinants of restenosis were quickly recognized; early candidate gene studies demonstrated that a variety of SNPs in inflammatory mediators were associated with stent restenosis.²⁷

To prevent stent restenosis, antiplatelet therapy, most effectively with aspirin and clopidogrel therapy, has become the standard of care after percutaneous stent placement. Antiplatelet agents are required for a much longer time with drug-eluting stents than with bare metal stents, although the optimal timeframe for antiplatelet therapy after drug-eluting stent placement remains to be determined. Despite dual therapy, restenosis occurs in up to 25% of patients who have received stents;²⁸ at 6 months' follow-up, the Low Responsiveness to Clopidogrel and Sirolimus- or Paclitaxel-Eluting Stent Thrombosis (RE-CLOSE) trial demonstrated a restenosis rate of 3.1%.²⁹ This high rate was 3-fold higher in nonresponders to clopidogrel and 5-fold higher in nonresponders to aspirin.³⁰ At higher loading doses, rates of clopidogrel and aspirin resistance decline, but they are still associated with a high incidence of thrombosis and mortality.³⁰ Moreover, patients who demonstrate clopidogrel resistance also exhibit

higher rates of concurrent aspirin resistance, and this suggests shared mechanisms of resistance. Furthermore, women and African Americans are much more likely to demonstrate resistance to antiplatelet therapy than men and Caucasians, respectively.³¹ The CYP2C9 enzyme participates in the activation process of clopidogrel; one *CYP2C9*2* allele, a loss-of-function variant (causing reduced amounts or activity of the gene product), is associated with dual nonresponsiveness to both aspirin and clopidogrel and higher cardiac mortality from drug-eluting stent thrombosis.³² Other SNPs that have been proposed to lead to clopidogrel and/or aspirin resistance include genes encoding the platelet P2Y₁₂ receptor that binds the active metabolite of clopidogrel; glycoproteins Ia, Ib-a, IIIa, and VI; cyclooxygenases; P-selectin; and other CYP450 genes.^{32–34} The potential clinical utility of testing for this variant is illustrated by a study from the Institut National de la Santé et de la Recherche Médicale, which found that among young patients who survived their first MI and were placed on aspirin and clopidogrel therapy, those with the *CYP2C9*2* variant suffered significantly more recurrent coronary events in both acute and long-term follow-up.³⁵

Warfarin provides a key demonstration of the potential for pharmacogenetics. As a medication used in more than 2 million patients with a very narrow therapeutic index and substantial interindividual dosage variability, the pharmacokinetics and pharmacodynamics of the drug are of particular interest to ensure that anticoagulated patients receive effective treatment and avoid serious adverse events. The state of the art for warfarin monitoring is repeated international normalized ratio measurements with dedicated inpatient or outpatient teams, but in cases of diminished access or patient noncompliance, a method to better estimate the starting dose is desirable. Two genes—*CYP2C9* in the hepatic CYP450 metabolizing system and *VKORC1*, which encodes the enzyme that reduces vitamin K epoxide to active vitamin K—account for 30% of the warfarin dose variability; when they are added to other physical parameters, 79% of the total variability is approached.³⁶ Multicenter trials are underway to assess the clinical value of incorporating genotype data prospectively into warfarin dosing; preliminary data suggest that the benefit will be greatest in those patients with multiple slow-metabolizer variants (thus requiring substantially lower doses) or in outpatient settings without access to careful anticoagulation monitoring.^{37,38} There are now online resources (eg, www.warfarindosing.org) for dose calculation that include demographic, medical history, and genotype data to prospectively estimate individualized warfarin starting doses, and

it may become more commonplace to include genotyping in the initial assessment of patients requiring warfarin therapy.

Diagnostic and Prognostic Genomic Profiles

Diagnostic and prognostic genetic signatures are being developed to better identify cardiovascular disease entities and select the optimal therapeutic and management options. Sets of variants—either

Diagnostic and prognostic genetic signatures are being developed to better identify cardiovascular disease entities and select the optimal therapeutic and management options.

SNPs or known mutations in specific genes—are correlated with disease states to produce a high-throughput chip or array. Studies have been conducted to retrospectively associate disease prognosis, especially in the realm of oncology (recurrence versus remission, aggressive versus localized) and for cardiovascular outcomes such as heart failure progression and heart transplant success. The chips employ a genetic signature—a gene expression profile (what RNAs are being produced and translated into a protein) or a gene mutation panel—established in previous trials to have a statistically significant association with a particular outcome.

Although the original goal of gene expression profiles may have been to study the global patterns of gene activity and interaction for basic biological research purposes, there are evident applications of the findings for improved diagnosis, categorization, and predicted course of disease. For example, CardioChip, a microarray developed at Harvard with 10,848 heart-specific complementary DNAs (representing the transcribed or expressed genes), was first used for basic research in end-stage cardiomyopathy.³⁹ This chip-based approach, interrogating multiple genes at once for mutation, may have value as a diagnostic assay to rule out multiple inherited cardiomyopathies with 1 test, and Harvard now offers modified microarrays for clinical genetic testing in hypertrophic and dilated cardiomyopathy.

Combined academic and industry efforts are developing new cardiovascular genomic profiles,

which are regulated as in vitro medical device assays, to provide gene expression and mutation information to better guide therapy choice and management.

Combined academic and industry efforts are developing new cardiovascular genomic profiles, which are regulated as in vitro medical device assays, to provide gene expression and mutation information to better guide therapy choice and management.

The assays are developed by testing genomic platforms against patient tissues retrospectively, with significant associations sought between gene expression or mutation patterns and specific disease states. In this manner, an assay can establish analytical and clinical validity as a predictive tool. For instance, the start-up CardioDx, one of a handful of new genomic profiling companies, is developing gene signature tools designed to improve diagnostic and prognostic accuracy in CAD, arrhythmias, and heart failure.⁴⁰ We are now witnessing the first prospective clinical trials to evaluate the tests' clinical utility, assessing the costs and benefits of incorporating these genomic data into actual patient management.

SUMMARY

Although the driving force behind GWASs was originally the discovery of novel biological pathways through the use of scalable genome technology and its capacity for generating hypothesis-independent genetic associations, the results have generated much excitement about the potential clinical applications of these genetic markers for disease prediction, prevention, and diagnosis. Direct-to-consumer genetic testing companies now offer mutation analysis and SNP microarray sequencing based on the findings of traditional genetic research and GWASs before the clinical validity or utility of population screening has been established. It is important for physicians to understand the purpose and limitations of GWAS information when patients inquire about these services or how genetic testing can enhance their care.

The medical literature and public media have focused our attention on the promise of genomics, and cardiology is uniquely primed for the advances in predictive accuracy and intervention efficacy that genomic profiles and pharmacogenetics may bring. Genotyping to predict the metabolizing capacity for medications, especially those with narrow therapeutic indices (like many used in cardiovascular management), is now an important consideration for drug choice and dosage and for the prevention of adverse outcomes. Additionally, complex patients, such as the one presented previously, may be excellent candidates for novel therapies such as the newly Food and Drug Administration–approved thienopyridine platelet inhibitor, prasugrel, which is thought to bypass the CYP2C19 metabolic pathway and has been shown to decrease rates of in-stent thrombosis in comparison with clopidogrel, although it may result in higher rates of bleeding complications.⁴¹

The determination of the clinical validity and utility of genomic associations, pharmacogenetic assays, and genomic profiles is occurring in academic, controlled, prospective clinical trials and in industry postmarketing analysis of tests already clinically available. Whether predisposition and predictive data generated by retrospective laboratory studies have medical value will be determined over the coming years as genomic technology, utilization, and education advance. Questions about whether it will be helpful to screen patients carrying markers associated with abdominal or intracranial aortic aneurysms before a nearly fatal late-stage presentation and whether patients beginning antiplatelet or anticoagulation therapy should have a priori genotyping will begin to be answered. As we discern more genomic markers—SNPs, copy number variants, and rare alleles—that influence the development and course of cardiovascular disease, our goal as evidence-based clinicians will be to apply this knowledge judiciously and responsibly for more personalized and cost-effective care.

As we discern more genomic markers—SNPs, copy number variants, and rare alleles—that influence the development and course of cardiovascular disease, our goal as evidence-based clinicians will be to apply this knowledge judiciously and

responsibly for more personalized and cost-effective care.

DISCLOSURES

Potential conflict of interest: Nothing to report.

REFERENCES

1. Albert CM, Cook NR, Gaziano JM, et al. Effect of folic acid and B vitamins on risk of cardiovascular events and total mortality among women at high risk for cardiovascular disease: a randomized trial. *JAMA* 2008; 299: 2027–2036.
2. Biselli PM, Guerzoni AR, de Godoy MF, et al. Genetic polymorphisms involved in folate metabolism and concentrations of methylmalonic acid and folate on plasma homocysteine and risk of coronary artery disease. *J Thromb Thrombolysis*. doi://10.1007/s11239-009-0321-7. Available at: <http://www.springerlink.com>.
3. Cohen JC, Boerwinkle E, Mosley Jr, et al. Sequence variations in PCSK9, low LDL, and protection against coronary artery disease. *N Engl J Med* 2006; 354: 1264–1272.
4. Frikke-Schmidt R, Nordestgaard BG, Stene MCA, et al. Association of loss-of-function mutations in the ABCA1 gene with high-density lipoprotein cholesterol levels and risk of ischemic heart disease. *JAMA* 2008; 299: 2525–2531.
5. Liu S, Schmitz C, Stampfer MJ, et al. A prospective study of TaqIB polymorphism in the gene coding for cholesteryl ester transfer protein and risk of myocardial infarction in middle-aged men. *Atherosclerosis* 2002; 161: 469–474.
6. Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009; 41: 334–341.
7. Roden DM. Clinical practice. Long-QT syndrome. *N Engl J Med* 2008; 358: 169–176.
8. Splawski I, Timothy KW, Tateyama M. Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia. *Science* 2002; 297: 1333–1336.
9. Burke A, Creighton W, Mont E. Role of SCN5A Y1102 polymorphism in sudden cardiac death in blacks. *Circulation* 2005; 112: 798–802.
10. Prutkin JM, Sotoodehnia N. Genetics of sudden cardiac arrest. *Prog Cardiovasc Dis* 2008; 50: 390–403.
11. Makita N. Phenotypic overlap cardiac sodium channelopathies: individual or mutation-specific? *Circulation* 2009; 73: 810–817.
12. Hirschhorn JN. Genomewide association studies—illuminating biologic pathways. *N Engl J Med* 2009; 360: 1699–1701.
13. Goldstein DB. Common genetic variation and human traits. *N Engl J Med* 2009; 360: 1696–1698.
14. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007; 316: 1491–1493.
15. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007; 316: 1488–1491.
16. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; 357: 443–453.
17. Larson MG, Atwood LD, Benjamin EJ, et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med Genet* 2007; 8(suppl 1): S5.
18. Schunkert H, Götz A, Braund P, et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* 2008; 117: 1675–1684.
19. Anderson JL, Horne BD, Kolek MJ, et al. Genetic variation at the 9p21 locus predicts angiographic coronary artery disease prevalence but not extent and has clinical utility. *Am Heart J* 2008; 156: 1155–1162.
20. Chen SN, Ballantyne CM, Gotto AM Jr, et al. The 9p21 susceptibility locus for coronary artery disease and the severity of coronary atherosclerosis. *BMC Cardiovasc Disord* 2009; 9: 3.
21. Ye S, Willeit J, Kronenberg F, et al. Association of genetic variation on chromosome 9p21 with susceptibility and progression of atherosclerosis: a population-based, prospective study. *J Am Coll Cardiol* 2008; 52: 378–384.
22. Gudbjartsson DF, Arnar DO, Helgadottir A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* 2007; 448: 353–357.
23. Kääb S, Darbar D, van Noord C, et al. Large scale replication and meta-analysis of variants on chromosome 4q25 associated with atrial fibrillation. *Eur Heart J* 2009; 30: 813–819.
24. Juang JM, Chern YR, Tsai CT, et al. The association of human connexin 40 genetic polymorphisms with atrial fibrillation. *Int J Cardiol* 2007; 116: 107–112.
25. Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet*. doi://10.1038/ng.361. Available at: <http://www.nature.com>.
26. Marchiori A, Mosena L, Prins MH, et al. The risk of recurrent venous thromboembolism among heterozygous carriers of factor V Leiden or prothrombin G20210A mutation. A systematic review of prospective studies. *Haematologica* 2007; 92: 1107–1114.
27. Monraats PS, Pires NM, Agema WR, et al. Genetic inflammatory factors predict restenosis after percutaneous coronary interventions. *Circulation* 2005; 112: 2417–2425.
28. Snoep JD, Hovens MM, Eikenboom JC, et al. Clopidogrel nonresponsiveness in patients undergoing percutaneous coronary intervention with stenting: a systematic review and meta-analysis. *Am Heart J* 2007; 154: 221–231.
29. Buonamici P, Marcucci R, Migliorini A, et al. Impact of platelet reactivity after clopidogrel administration on drug-eluting stent thrombosis. *J Am Coll Cardiol* 2007; 49: 2312–2317.
30. Gori AM, Marcucci R, Migliorini A, et al. Incidence and clinical impact of dual nonresponsiveness to aspirin and clopidogrel in patients with drug-eluting stents. *J Am Coll Cardiol* 2008; 52: 734–739.
31. Gurbel PA, Bliden KP, Cohen E, et al. Race and sex differences in thrombogenicity: risk of ischemic events

- following coronary stenting. *Blood Coagul Fibrinolysis* 2008; 19: 268–275.
32. Giusti B, Gori AM, Marcucci R, et al. Relation of cytochrome P450 2C19 loss-of-function polymorphism to occurrence of drug-eluting coronary stent thrombosis. *Am J Cardiol* 2009; 103: 806–811.
33. Nguyen TA, Diodati JG, Pharand C. Resistance to clopidogrel: a review of the evidence. *J Am Coll Cardiol* 2005; 45: 1157–1164.
34. Szczeklik A, Muisal J, Undas A, et al. Aspirin resistance. *J Thromb Haemost* 2005; 3: 1655–1662.
35. Collet JP, Hulot JS, Pena A, et al. Cytochrome P450 2C19 polymorphism in young patients treated with clopidogrel after myocardial infarction: a cohort study. *Lancet* 2009; 373: 309–317.
36. Daly AK. Pharmacogenomics of anticoagulants: steps toward personal dosage. *Genome Med* 2009; 1: 10.
37. Anderson JL, Horne BD, Stevens SM, et al. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. *Circulation* 2007; 116: 2563–2570.
38. Scott SA, Edelmann L, Kornreich R, et al. Warfarin pharmacogenetics: CYP2C9 and VKORC1 genotypes predict different sensitivity and resistance frequencies in the Ashkenazi and Sephardi Jewish populations. *Am J Hum Genet* 2008; 82: 495–500.
39. Barrans JD, Allen PD, Stamatiou D, et al. Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* 2002; 160: 2035–2043.
40. Sehnert AJ, Daniels SE, Elashoff M, et al. Lack of association between adrenergic receptor genotypes and survival in heart failure patients treated with carvedilol or metoprolol. *J Am Coll Cardiol* 2008; 52: 644–651.
41. Wiviott SD, Braunwald E, McCabe CH, et al. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med* 2007; 357: 2001–2015.



Statement of Ownership, Management, and Circulation
(All Periodicals Publications Except Requester Publications)

1 Publication Title
Mount Sinai Journal of Medicine

2 Issue Frequency
6x's/yr in April, June, Aug, Sept, Nov, Dec

3 Issue Date for Circulation Data Below
August 2009

4 Number of Copies
5048

5 Total Number of Copies (Net press run)
5048

6 Paid Circulation Outside the Mail
0

7 Paid Distribution Outside the Mail
0

8 Paid Distribution by Other Classes of Mail Through the USPS (e.g. First-Class Mail®)
0

9 Total Paid Distribution (Sum of 6, 7, and 8)
0

10 Free or Nominal Rate Outside-County Copies Included on PS Form 3541
377

11 Free or Nominal Rate In-County Copies Included on PS Form 3541
0

12 Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g. First-Class Mail)
0

13 Free or Nominal Rate Distribution Outside the Mail (Carriers or other means)
0

14 Total Free or Nominal Rate Distribution (Sum of 10, 11, 12, and 13)
377

15 Total Distribution (Sum of 14 and 9)
3005

16 Copies not Distributed (See Instructions 1-4)
2043

17 Total (Sum of 15 and 16)
5048

18 Percent Paid
87.49

19 Publication of Statement of Ownership
If the publication is a general publication, publication of this statement is required. Will be printed in the December 2009 issue of this publication.

20 Signature and Title of Editor, Publisher, Business Manager, or Owner
Elizabeth Konkle, Associate Financial Manager

21 Date
10/1/09



Discover
Something
Great

Distinctive Content

Content

*Journals • Backfile Collections
Reference Works & Encyclopedias
OnlineBooks™
Current Protocols Lab Manuals
Evidence-Based Resources*

Access

*Single article/chapter or full title
One-time fee, subscription,
or Pay-Per-View options
Unlimited concurrent usage*

Capabilities

*E-mail Alerts • EarlyView®
Citation download & tracking
Remote access • Mobile Edition®*

Promote the process of scientific discovery with a different breed of research system: **Wiley InterScience®**. Long distinguished by its scope and extensive coverage of emerging disciplines, Wiley's authoritative content spans the spectrum of scientific, medical, technical, and business specialties. The dynamic **Wiley InterScience** platform adds value to your research by providing simple desktop access to this preeminent digital library of journals, reference works, books, laboratory manuals, evidence-based databases, and more. Researchers reap the combined benefits of top-tier resources and up-to-the-minute functionality. Experience the difference **Wiley InterScience** can make at your institution today.



1807-2007 KNOWLEDGE FOR GENERATIONS

www.interscience.wiley.com

NOTHING BEATS BEING FIRST

SIGN UP FOR FREE

Journal of Hospital Medicine

e-alerts and be the first to know
about new research in your field!

It's fast, free, and flexible.
Customize your alerts to deliver
only the information you need.

- **Table of Contents Alerts:**

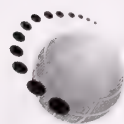
Get notifications when new issues
have published.

- **Automated Search Alerts:**

Retrieve and re-run detailed search queries
and set a Saved Search e-mail Alert, which
automatically notifies you when new content
that matches the specified search comes online.

- **Citation Alerts:**

Get notification when an article has been cited.



WILEY

InterScience®

DISCOVER SOMETHING GREAT



WILEY-
BLACKWELL

shm
Society of Hospital Medicine

www.JournalofHospitalMedicine.com

NOTHING BEATS BEING FIRST

SIGN UP FOR FREE

Journal of Hospital Medicine
e-alerts and be the first to know
about new research in your field!

It's fast, free, and flexible.
Customize your alerts to deliver
only the information you need.

- **Table of Contents Alerts:**

Get notifications when new issues
have published.

- **Automated Search Alerts:**

Retrieve and re-run detailed search queries
and set a Saved Search e-mail Alert, which
automatically notifies you when new content
that matches the specified search comes online.

- **Citation Alerts:**

Get notification when an article has been cited.

A dark blue banner with a white hand icon pointing upwards on the left. The hand has the text 'SIGN UP TODAY!' on it. To the right of the hand are the logos for 'WILEY-BLACKWELL' and 'shm Society of Hospital Medicine'. At the bottom of the banner, the website URL 'www.JournalofHospitalMedicine.com' is displayed in white.

SIGN UP TODAY!

WILEY-BLACKWELL

shm
Society of Hospital Medicine

www.JournalofHospitalMedicine.com



HINARI PHASE I COUNTRIES

Free access to more than 100 biomedical journals published by Wiley

HINARI PHASE II COUNTRIES

Nominal cost access to more than 100 biomedical journals published by Wiley

Developing World Access to Research

“perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries.”

Former WHO Director-General Dr Gro Harlem Brundtland (July, 2001)

Wiley, together with the World Health Organisation and fellow Scientific, Technical and Medical publishers, is proud to play a role in furthering global dissemination of biomedical information.

Through the Health InterNetwork Access to Research Initiative (HINARI) www.healthinternetwork.org, free or nominal cost online access to more than 2,200 biomedical journals from 42 publishers is provided to more than 1,000 institutions in approximately 100 of the world's lowest income countries.

Wiley is one of the six founding publishers who led this initiative with the WHO. We offer more than 100 of our leading biomedical journals through the HINARI program; including many leading journals published with our society partners.

HINARI
ACCESS TO RESEARCH



WILEY
InterScience®
DISCOVER SOMETHING GREAT
www.interscience.wiley.com

Wiley – helping to make a world of difference.



Fresh Content Sliced Any Way You Like It



Wiley InterScience® OnlineBooks™

Content

*Thousands of Wiley book titles
across science, technology,
medicine, business, and finance*

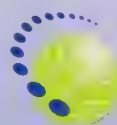
Access

*Single chapter or full title
One-time fee, subscription,
or Pay-Per-View options*

Capabilities

*Search at the chapter level
Full text presented in PDF format
Citation download feature
COUNTER-compliant usage statistics*

Satisfy your library's evolving research needs and budget with a customized blend of new content and licensing options at **Wiley InterScience® OnlineBooks™**. Pick and choose from an expanding roster of thousands of Wiley titles available online via **Wiley InterScience** to create a dynamic electronic book collection. Select from our new flexible licensing models – one-time fee, subscription, or Pay-Per-View. Add books as you need them from our distinguished portfolio of classic and cutting-edge titles. Whichever combination you select, enjoy unlimited concurrent access to content across the spectrum of science, medicine, technology, business, and finance.



WILEY
InterScience®
DISCOVER SOMETHING GREAT



1807-2007 KNOWLEDGE FOR GENERATIONS

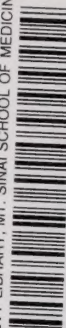
www.interscience.wiley.com







LEVY LIBRARY, MT. SINAI SCHOOL OF MEDICINE



3 4805 0717666 8